TRANSACTIONS
OF THE
AMERICAN INSTITUTE
OF
ELECTRICAL ENGINEERS

VOLUME 75
1956
PART I. COMMUNICATION AND ELECTRONICS

PUBLISHED BY THE
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
33 WEST 39TH STREET
NEW YORK 18, N. Y., U. S. A.

Copyright 1957 by the American Institute of Electrical Engineers
Printed in U. S. A. by The Mack Printing Company, Easton, Pa.

Preface

The AIEE *Transactions* for 1956 (volume 75) is published in three parts: Part I. Communication and Electronics; Part II. Applications and Industry; and Part III. Power Apparatus and Systems. The papers in each of the three parts are classified according to subject matter as follows:

Part I. Communication and Electronics

- Communication Switching Systems
- Communication Theory
- Radio Communications Systems
- Special Communications Applications
- Telegraph Systems
- Television and Aural Broadcasting Systems
- Wire Communications Systems
- Basic Sciences
- Computing Devices
- Dielectrics
- Electrical Techniques in Medicine and Biology
- Electronics
- Instruments and Measurements
- Magnetic Amplifiers
- Metallic Rectifiers
- Nucleonics
- Solid State Devices

Part II. Applications and Industry

- Air Transportation
- Domestic and Commercial Applications
- Land Transportation
- Marine Transportation
- Production and Application of Light
- Chemical Industry
- Electric Heating
- Electric Welding
- Feedback Control Systems
- General Industry Applications
- Industrial Control
- Industrial Power Rectifiers
- Industrial Power Systems
- Mining and Metal Industry
- Petroleum Industry

Part III. Power Apparatus and Systems

- Carrier Current
- Insulated Conductors
- Power Generation
- Protective Devices
- Relays
- Rotating Machinery
- Substations
- Switchgear
- System Engineering
- Transformers
- Transmission and Distribution

- Education
- Safety
- Research

Each part has been indexed separately in the back of that particular part. The three parts are not cross-referenced; hence the user should determine first whether the subject matter of the paper desired is in the field of communication and electronics, applications and industry, or power apparatus and systems. Papers are listed in the subject index under several key words in the titles. The original numbers assigned to the papers are given in the author index. Volume 75 contains the technical papers and related discussions presented at these meetings:

1. Winter General Meeting, New York, N. Y., January 30–February 3, 1956.
2. South West District Meeting, Dallas, Tex., April 2–4, 1956.
3. Great Lakes District Meeting, Fort Wayne, Ind., April 16–18, 1956.
4. North Eastern District Meeting, Rochester, N. Y., May 2–4, 1956.
5. Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956.
6. Fall General Meeting, Chicago, Ill., October 1–5, 1956.

Statements and opinions given in the papers and discussions published in *Transactions* are the expressions of the contributors for which the American Institute of Electrical Engineers assumes no responsibility.

Communication and Electronics Index for 1956

1. Technical Subject Index

A2A Video System, Field Experience with the. Gast. 710–16
A-C Bridge to Measure Core Loss at High Inductions, The Use of an. Cooter, Harris....252–5; disc. 256
A-C Kelvin Bridge for the Audio-Frequency Range, An. Dunfee....123–7; disc. 127
Accuracy Obtainable with Multiplier Phototubes, Synchronization. Levi....603–06
Accurate Determination of the Capacitance of a Thin Rectangular Plate. Reitan, Higgins....761–6; disc. 766
Air-Dielectric Cable, Styroflex Aluminum-Sheathed. Merrell, McKean, Arbuthnott...669–74; disc. 674
Air Purification, Heat as a Means of. Beckett, Clifton. 373–7
Alternating Current Control of the Half-Wave Bridge Magnetic Amplifier. Bernstein, Schmitz....433–8
Aluminum-Sheathed Air-Dielectric Cable, Styroflex. Merrell, McKean, Arbuthnott...669–74; disc. 674
Amplifiers, Application of Junction Transistors to Carrier-Frequency Computing. Curtin....746–52; disc. 752
Amplifiers for Flight Simulators, Transistor Analogue Computing. Weyrick....338–42
Amplifiers, Hysteresis Loops in Dielectric. Wingrove, Depian, Shevel....283–8; disc. 288
Amplifiers, Optimum Design of Common Emitter Transistor Audio. Vallese....391–5
Amplifiers, Temperature Stabilization of Transistor. Vallese....379–84
Amplifiers with Switched Mode of Operation, Transistor Power. Milnes....368–72
Amplifiers with 2-Phase Motor Load, Measurements on High-Speed Magnetic Servo. Geyger.......242–8
Analogue Computation with Magnetic Cores, Transcendental Function. Schaefer, Van Allen....160–4; disc. 164
Analogue Computer Generation of Probability Distributions for Operations Research. Diamantides.. 86–91
Analogue Computer, Linear Programming on an Electronic. Pyne....139–42; disc. 142
Analogue Computer Synthesis and Error Matrices. Honnell, Horn....26–32
Analogue Computing Amplifiers for Flight Simulators, Transistor. Weyrick....338–42
Analogue to Digital Converters, Cyclic Decimal Codes for. O'Brien....120–2
Analysis, A Phasor Method of Nonlinear Network. Neal....630–6; disc. 636
Analysis and Design of the Ferroelectric Resonant Trigger Pair, An. Gremer....404–07
Analysis and Design, The Ferroresonant Trigger Pair:. Spitzer....407–16
Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop Core Material in Series with Capacitors. Salihi..296–305; disc. 305
Analysis of Junction Transistor Flip-Flops, D-C Graphical. Bashkow....1–7
Analysis of Linear Sampled-Data Systems with Finite Pulse Width: Open Loop. Farmanfarma..808–19
Analysis of Sampled Data Servomechanisms Performed on the IBM Type 650, The. Tostanoski....446–50
Analysis with a Digital Computation System, Automatic Network. Wong, Kochen........172–5; disc. 176
Analytical Design and Evaluation of Electromagnets, The. Kelly, Wallace....675–80
Antennas, Impedance of Thin-Wire Loop. Storer.... 606–18
Application of a Digital Computer to the Design of Power Transformers to Specification. Moore, Duboc, Zaphyr....134–7; disc. 137
Application of Junction Transistors to Carrier-Frequency Computing Amplifiers. Curtin...746–52; disc. 752
Application of Statistical Techniques to Electron Tubes for Use in a 4,000 Mile Transmission System. Van Haste....50–4
Application of the Type-P1 Carrier System to Rural Telephone Lines. Howard......205–13; disc. 214
Approximation Problem, Note on the. Balabanian.72–5
Arc in a Magnetic Field, Properties of a D-C. Winsor, Lee....143–8; disc. 148
Arc-Resistant Molding Materials and Finishes. Sterling 776–80
Atlantic Link, Power-Feed Equipment for the North. Meszaros, Spencer....890–8
Atlantic Link, Repeater Design for the North. Gleichmann, Lince, Wooley, Braga....865–77
Atlantic Link, Repeater Production for the North. Lamb, Heffner....877–89
Atlantic Link, System Design for the North. Lewis, Tucker, Lovell, Fraser....850–64
Atomic-Wave Functions, A Digital Computer Program for Determining. Piper....152–5
Audio Amplifiers, Optimum Design of Common-Emitter Transistor. Vallese....391–5
Audio-Frequency Range, An A-C Kelvin Bridge for the. Dunfee....123–7; disc. 127
Automatic Dispatch System for Half-Duplex Teletypewriter Lines. Maxwell, Farrar....441–5
Automatic Manufacturing Facilities for Reed Switches, The Development of. Hosford....496–500
Automatic Network Analysis with a Digital Computation System. Wong, Kochen....172–5; disc. 176
Automatic Protection Switching for TD-2 Radio System. Klemm....520–7
Axis into Two Parallel R-C Ladders and an Ideal Transformer, Synthesis of Transfer Functions with Poles Restricted to the Negative Real. Malti, Sun..... ....165–71; disc. 171

B

Barium Titanate Ceramics, Investigations Concerning Polarization in. Marks, Waidelich, Monson.... 469–77
(Basic Sciences) A New Method for Treating Nonlinear Problems with Applications to Iron-Cored Oscillatory Circuits. Gohar:
  Part I. Subharmonics Do Not Exist.........771–6
(Basic Sciences) A Phasor Method of Nonlinear Network Analysis. Neal....630–6; disc. 636
(Basic Sciences) Accurate Determination of the Capacitance of a Thin Rectangular Plate. Reitan, Higgins....761–6; disc. 766
(Basic Sciences) An Analysis and Design of the Ferroelectric Resonant Trigger Pair. Gremer....404–07
(Basic Sciences) Analogue Computer Synthesis and Error Matrices. Honnell, Horn....26–32
(Basic Sciences) Analysis of Linear Sampled-Data Systems with Finite Pulse Width: Open Loop. Farmanfarma....808–19
(Basic Sciences) Computation of the Impedances of Nonuniform Lines by a Direct Method. Pipes 551–4
(Basic Sciences) Dynamax, a New Crystal and Domain-Oriented Magnetic Core Material. Howe....... ....548–50; disc. 551
(Basic Sciences) Electrical Formulas in Inch Units of Length. Williamson....61–4
(Basic Sciences) Electrical Units and Dimensions. Young....767–8; disc. 768
(Basic Sciences) Graphical Determination of the Fourier Series Coefficients. Middendorf....478–82
(Basic Sciences) Integrated s-Plane Synthesis Using 2-Way Root Locus. Zaborszky....797–801
(Basic Sciences) Investigations Concerning Polarization in Barium Titanate Ceramics. Marks, Waidelich, Monson....469–77
(Basic Sciences) Magnetic Effects of Compressional Stress at Low Field Intensities. Fischell........ ....148–51; disc. 151
(Basic Sciences) Note on the Approximation Problem. Balabanian....72–5
(Basic Sciences) Properties of a D-C Arc in a Magnetic Field. Winsor, Lee....143–8; disc. 148
(Basic Sciences) Representation of Nonlinear Characteristics by Linear Elements. Abou-Hussein..84–6
(Basic Sciences) Some Mathematical Properties of Root Loci for Control System Design. Reza.....103–08
(Basic Sciences) Subharmonic Response of the Ferroresonant Circuit with Coil Hysteresis. Brenner... 450–6
(Basic Sciences) Superposed Magnetic Fields in Materials with Rectangular Hysteresis Loops. Wakeman, Beck....562–9
(Basic Sciences) Synthesis of Transfer Functions with Poles Restricted to the Negative Real Axis into Two Parallel R-C Ladders and an Ideal Transformer. Malti, Sun....165–71; disc. 171
(Basic Sciences) The Analytical Design and Evaluation of Electromagnets. Kelly, Wallace........675–80
(Basic Sciences) The Ferroresonant Trigger Pair: Analysis and Design. Spitzer....407–16
(Basic Sciences) The Impedance Tensor of the General Machine. Yu....181–7
Behavior of a 3-Phase Magnetic Amplifier, Dynamic. Ellert....554–60; disc. 560
Bell System Teletypewriters, Static Elimination on. Bavley, Perry, Swesey....92–4
Bell System Wire Spring Relays, Design Features of. Knapp....482–6
Bilateral Switching Element, The Symmetrical Transistor as a. Trousdale....400–03
Black and White to Color Television, Conversion of Studio Lighting from. Wygant....75–8
Bonneville Power Administration, Southwest Oregon Radio System for the. Marihart, Wylie....531–8
Bridge Magnetic Amplifier, Alternating Current Control of the Half-Wave. Bernstein, Schmitz......433–8
Bridge Magnetic Amplifiers with Inductive Loads, Full-Wave. Woodson....7–9
Bridge to Measure Core Loss at High Inductions, The Use of an A-C. Cooter, Harris....252–5; disc. 256
Built-In Reliability, Microwave Facilities with. Kuck 438–41
Business Machine Data Over Standard Telegraph Channels, Transmission of. Bramhall.416–20; disc. 420

C

Cable Design and Manufacture for the Transatlantic Submarine Cable System. Lebert, Fischer, Biskeborn....907–17
Cable Laying for the Transatlantic Cable System, Route Selection and. Jack, Leech, Lewis...944–55
Cable, Lightning Protection on the Stevens Point–Wisconsin Rapids Intercity Telephone. Dowling.. 697–701
(Cable) Power-Feed Equipment for the North Atlantic Link. Meszaros, Spencer....890–8
Cable) Power-Feed System for the Newfoundland–Nova Scotia Link. Thomas, Kelly........938–43
(Cable) Repeater Design for the Newfoundland–Nova Scotia Link. Brockbank, Walker, Welsby..927–38
(Cable) Repeater Design for the North Atlantic Link. Gleichmann, Lince, Wooley, Braga.........865–77
(Cable) Repeater Production for the North Atlantic Link. Lamb, Heffner....877–89
Cable, Styroflex Aluminum-Sheathed Air-Dielectric. Merrell, McKean, Arbuthnott....669–74; disc. 674
(Cable) System Design for the Newfoundland–Nova Scotia Link. Halsey, Bampton....917–27
(Cable) System Design for the North Atlantic Link. Lewis, Tucker, Lovell, Fraser....850–64
Cable System, Electron Tubes for the Transatlantic. McNally, Metson, Veazie, Holmes........898–907
Cable System—Planning and Overall Performance, Transatlantic Telephone. Mottram, Halsey, Emling, Griffith....842–50
(Cable) Transatlantic Communications: A Historical Résumé. Kelly, Radley....841–2
Calculations with Digital Computers, Transformer Reactance. Rabins....261–7; disc. 267
Calculator, A New Five-Digit Plugboard System for Card-Programmed. Williams, Noonan.....67–71
Canada, Development of Transcontinental Communications in. Noyes, Gaudet, Bonneville.....342–52
Capability of Sealed Contact Relays. Hovgaard..466–8
Capacitance of a Thin Rectangular Plate, Accurate Determination of the. Reitan, Higgins....761–6; disc. 766
Capacitors, Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop Core Material in Series with. Salihi..296–305; disc. 305
Capacitors, Signals from Switched Ferroelectric Memory. Pulvari, McDuffie....681–5
Card-Programmed Calculator, A New Five-Digit Plugboard System for. Williams, Noonan.......67–71
Carrier-Frequency Computing Amplifiers, Application of Junction Transistors to. Curtin....746–52; disc. 752
Carrier Frequency Networks, Telephone. Wallenstein 691–7
Carrier System, Circuit and Equipment Descriptions of Type-P1. Eberhart, Hallenbeck, Perkins..195–204
Carrier System Employing Transistors, A New 4-Channel Open-Wire. Caruthers, Chaskin, Krengel..314–22
Carrier System, Objectives and General Description of the Type-P1. Boyd....188–91
Carrier System, Power Supplies for the Type-P1. Smith ....191–4; disc. 194
Carrier System to Rural Telephone Lines, Application of the Type-P1. Howard.......205–13; disc. 214

Cathode-Ray Tube, Method of Measuring Spot Size of. Kuhn, Levine....357–9; disc. 359
Ceramics, Investigations Concerning Polarization in Barium Titanate. Marks, Waidelich, Monson....469–77
Channels, Principles of Noise Reduction in Communication. Schwartz....44–50
Characteristics by Linear Elements, Representation of Nonlinear. Abou-Hussein....84–6
Characteristics of Toroidal Cores Used in Magnetic Amplifiers, Low-Impedance Operational. Boyajian....307–13
Circuit and Equipment Descriptions of Type-P1 Carrier System. Eberhart, Hallenbeck, Perkins...195–204
Circuit Control System Design Considerations, Magnetic Logic. Van Nice....595–600
Circuit; Its Operation and Mode Transition, The Simple Reactor. Ringelman, Fenaroli....660–8; disc. 668
Circuit Performance, Influence of Noise on Telephone Signaling. Weber....636–42
Circuit, Power Transistor Switching. Huang, Slobodzinski....290–6
Circuit, The Pulse Stretch Coupling. Patton....377–9
Circuit with Coil Hysteresis, Subharmonic Response of the Ferroresonant. Brenner....450–6
Circuits, A New Method for Treating Nonlinear Problems with Applications to Iron-Cored Oscillatory. Gohar:
  Part I. Subharmonics Do Not Exist....771–6
Circuits, Several Valued Combinational Switching. Lee, Chen....278–83
Codes for Analogue to Digital Converters, Cyclic Decimal. O'Brien....120–2
Coefficients, Graphical Determination of the Fourier Series. Middendorf....478–82
Coil Hysteresis, Subharmonic Response of the Ferroresonant Circuit with. Brenner....450–6
Color Television, Conversion of Studio Lighting from Black and White to. Wygant....75–8
Color-Television-System Performance Requirements. Kennedy....653–9
Combinational Switching Circuits, Several Valued. Lee, Chen....278–83
Communication Channels, Principles of Noise Reduction in. Schwartz....44–50
Communication Engineer's Needs in Information Theory, The. Rea....805–08
(Communication Switching Systems) Capability of Sealed Contact Relays. Hovgaard....466–8
(Communication Switching Systems) Design Features of Bell System Wire Spring Relays. Knapp....482–6
(Communication Switching Systems) Electronics in Telephone Switching Systems. Joel....701–09; disc. 710
Communication Switching Systems, Manufacture of Wire Spring Relays for. Rice....513–18
(Communication Switching Systems) Recent Developments in 4-Wire Switching of Long Distance Telephone Circuits. Molnar....722–32; disc. 732
(Communication Switching Systems) Recent New Features for the No. 5 Crossbar Switching System. Dehn, Hersey....457–66
(Communication Switching Systems) Several Valued Combinational Switching Circuits. Lee, Chen....278–83
(Communication Switching Systems) The Development of Automatic Manufacturing Facilities for Reed Switches. Hosford....496–500
(Communication Switching Systems) The Full Stature of the Crossbar Tandem Switching System. Meszar....486–96
(Communication Switching Systems) Traffic Considerations in Community Dial Offices. Lanigan, McDonough....543–7; disc. 547
(Communication Theory) A First Look at Random Noise. Rice....128–31
(Communication Theory) The Technical Feasibility of Translating Languages by Machine. Yngve.792–7
Communications: A Historical Résumé Transatlantic. Kelly, Radley....841–2
Communications in Canada, Development of Transcontinental. Noyes, Gaudet, Bonneville...342–52
Community Dial Offices, Traffic Considerations in. Lanigan, McDonough....543–7; disc. 547
Companies, Microwave for Telephone. Fisher..538–42
Comparison of Three Common Emitter Transistor Servo Preamplifiers, A. DeSautels....17–25
Components, The Use of Computing Machines in the Statistical Evaluation of Electrical. Andrus....501–05
Compressional Stress at Low Field Intensities, Magnetic Effects of. Fischell....148–51; disc. 151
Computation of the Impedances of Nonuniform Lines by a Direct Method. Pipes....551–4
Computation System, Automatic Network Analysis with a Digital. Wong, Kochen....172–5; disc. 176
Computation with Magnetic Cores, Transcendental Function Analogue. Schaefer, Van Allen....160–4; disc. 164
Computer—A University Research Project, The Wisconsin Integrally Synchronized. Asmuth, Davidson, Miller, Noble, Scidmore....330–8
Computer Generation of Probability Distributions for Operations Research, Analogue. Diamantides....86–91
Computer, Linear Programming on an Electronic Analogue. Pyne....139–42; disc. 142
Computer Program for Determining Atomic-Wave Functions, A Digital. Piper....152–5
Computer Synthesis and Error Matrices, Analogue. Honnell, Horn....26–32
Computer to the Design of Power Transformers to Specification, Application of a Digital. Moore, Duboc, Zaphyr....134–7; disc. 137
Computers, Transformer Reactance Calculations with Digital. Rabins....261–7; disc. 267
Computing Amplifiers, Application of Junction Transistors to Carrier-Frequency. Curtin....746–52; disc. 752
Computing Amplifiers for Flight Simulators, Transistor Analogue. Weyrick....338–42
(Computing Devices) A New Five-Digit Plugboard System for Card-Programmed Calculator. Williams, Noonan....67–71
(Computing Devices) Cyclic Decimal Codes for Analogue to Digital Converters. O'Brien....120–2
(Computing Devices) Magnetic Data Recording Theory: Head Design. Hoagland....506–12
(Computing Devices) Numerical Integration of Differential Equations on the 704 EDPM. Anderson, Johnson....569–74
(Computing Devices) Optimum Linear-Segment Function Generation. Hamer....518–20
(Computing Devices) Some of the Engineering Aspects of the Machine Translation of Language. Wall....580–4; disc. 585
(Computing Devices) The Analysis of Sampled Data Servomechanisms Performed on the IBM Type 650. Tostanoski....446–50
Computing Machines in the Statistical Evaluation of Electrical Components, The Use of. Andrus....501–05
Contact Relays, Capability of Sealed. Hovgaard.466–8
Control of the Half-Wave Bridge Magnetic Amplifier, Alternating Current. Bernstein, Schmitz....433–8
Control System Design Considerations, Magnetic Logic Circuit. Van Nice....595–600
Control System Design, Some Mathematical Properties of Root Loci for. Reza....103–08
Control System for Integrated Data Processing. Easterlin....421–7
Conversion of Studio Lighting from Black and White to Color Television. Wygant....75–8
Converters, Cyclic Decimal Codes for Analogue to Digital. O'Brien....120–2
Co-ordinated Protection for Open-Wire Joint Use—Minneapolis Tests. Albrecht, Dietz, Christoferson, Slothower....217–22; disc. 223
Co-ordinated Protection for Open-Wire Joint Use—Ontario Tests. Ellis, Phelps, Roach, Treen....223–34; disc. 235
Co-ordinated Protection for Open-Wire Joint Use—Present Trends. Bullard, Weppler....215–16
Core Loss at High Inductions, The Use of an A-C Bridge to Measure. Cooter, Harris....252–5; disc. 256
Core Material in Series with Capacitors, Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop. Salihi....296–305; disc. 305
Cores, Flux Reversal in Magnetic Amplifier. Friedlaender....268–76; disc. 277
Cores, Transcendental Function Analogue Computation with Magnetic. Schaefer, Van Allen....160–4; disc. 164
Cores Used in Magnetic Amplifiers, Low-Impedance Operational Characteristics of Toroidal. Boyajian....307–13
Corona Field Strength and Current Density in Electrical Precipitation, A New Technique for the Measurement of. Cooperman....64–7
Correction, Delay Distortion. Cannon....55–61
Coupling Circuit, The Pulse Stretch. Patton....377–9
Crossbar Switching System, Recent New Features for the No. 5. Dehn, Hersey....457–66
Crossbar Tandem Switching System, The Full Stature of the. Meszar....486–96
Crystal and Domain-Oriented Magnetic Core Material, Dynamax, a New. Howe....548–50; disc. 551
Current Density in Electrical Precipitation, A New Technique for the Measurement of Corona Field Strength and. Cooperman....64–7
Cyclic Decimal Codes for Analogue to Digital Converters. O'Brien....120–2

D

Data Over Standard Telegraph Channels, Transmission of Business Machine. Bramhall....416–20; disc. 420
Data Processing, Control System for Integrated. Easterlin....421–7
Data Processing, Simplified Printing Telegraph Switching and Integrated. Booth, Klich....11–14
Data Recording Theory: Head Design, Magnetic. Hoagland....506–12
Data Transmission Machine, A. Doty, Tate....600–03
D-C Arc in a Magnetic Field, Properties of a. Winsor, Lee....143–8; disc. 148
D-C Graphical Analysis of Junction Transistor Flip-Flops. Bashkow....1–7
Decimal Codes for Analogue to Digital Converters, Cyclic. O'Brien....120–2
Delay Distortion Correction. Cannon....55–61
Demand, Measuring 1,800,000-Kw. Morris, Thomas....95–7
Design and Evaluation of Electromagnets, The Analytical. Kelly, Wallace....675–80
Design Considerations, Magnetic Logic Circuit Control System. Van Nice....595–600
Design Features of Bell System Wire Spring Relays. Knapp....482–6
Design of Common-Emitter Transistor Audio Amplifiers, Optimum. Vallese....391–5
Design of Electronic Equipment for Radio-Interference Reduction. Albin, Sachs....352–7
Design of the Ferroelectric Resonant Trigger Pair, An Analysis and. Gremer....404–07
Design of Wide-Band R-F Transformers Utilizing a Synthesized Equivalent Network. Kajihara.802–05
Design, The Ferroresonant Trigger Pair: Analysis and. Spitzer....407–16
Determination of the Fourier Series Coefficients, Graphical. Middendorf....478–82
Development of Automatic Manufacturing Facilities for Reed Switches, The. Hosford....496–500
Development of Transcontinental Communications in Canada. Noyes, Gaudet, Bonneville....342–52
Developments in 4-Wire Switching of Long Distance Telephone Circuits, Recent. Molnar....722–32; disc. 732
Dial Offices, Traffic Considerations in Community. Lanigan, McDonough....543–7; disc. 547
Dielectric Amplifiers, Hysteresis Loops in. Wingrove, Depian, Shevel....283–8; disc. 288
(Dielectrics) A Field Survey of Transformer Oil Quality. Degnan, Shimanski....575–9; disc. 579
(Dielectrics) Arc-Resistant Molding Materials and Finishes. Sterling....776–80
(Dielectrics) Nontracking Organic Insulations. Norman, Pfuntner, Kessel....257–61
Differential Equations on the 704 EDPM, Numerical Integration of. Anderson, Johnson....569–74
Digital Computation System, Automatic Network Analysis with a. Wong, Kochen....172–5; disc. 176
Digital Computer Program for Determining Atomic-Wave Functions, A. Piper....152–5
Digital Computer to the Design of Power Transformers to Specification, Application of a. Moore, Duboc Zaphyr....134–7; disc. 137
Digital Computers, Transformer Reactance Calculations with. Rabins....261–7; disc. 267
Digital Converters, Cyclic Decimal Codes for Analogue to. O'Brien....120–2
Dimensions, Electrical Units and. Young....767–8; disc. 768
Diode-Shunting in Magnetic Amplifiers. Lowrance, Dolan....619–24
Direct Water-Cooled Germanium Power Rectifier. Wahl....832–41
Dispatch System for Half-Duplex Teletypewriter Lines, Automatic. Maxwell, Farrar....441–5
Distortion Correction, Delay. Cannon....55–61
Domain-Oriented Magnetic Core Material, Dynamax, a New Crystal and. Howe....548–50; disc. 551
Duplex Teletypewriter Lines, Automatic Dispatch System for Half-. Maxwell, Farrar....441–5
Dynamax, a New Crystal and Domain-Oriented Magnetic Core Material. Howe....548–50; disc. 551
Dynamic Behavior of a 3-Phase Magnetic Amplifier. Ellert....554–60; disc. 560

E

Electrical Components, The Use of Computing Machines in the Statistical Evaluation of. Andrus....501–05
Electrical Formulas in Inch Units of Length. Williamson....61–4
Electrical Precipitation, A New Technique for the Measurement of Corona Field Strength and Current Density in. Cooperman....64–7
Electrical Units and Dimensions. Young....767–8; disc. 768
Electromagnets, The Analytical Design and Evaluation of. Kelly, Wallace....675–80
Electron Tubes for the Transatlantic Cable System. McNally, Metson, Veazie, Holmes....898–907
Electron Tubes for Use in a 4,000 Mile Transmission System, Application of Statistical Techniques to. Van Haste....50–4
Electronic Equipment for Radio-Interference Reduction, Design of. Albin, Sachs....352–7
(Electronics) A Comparison of Three Common Emitter Transistor Servo Preamplifiers. DeSautels...17–25

Vol. 75
Pt. 1
1956

(Electronics) A New Technique for the Measurement of Corona Field Strength and Current Density in Electrical Precipitation. Cooperman........64–7
(Electronics) Application of Junction Transistors to Carrier-Frequency Computing Amplifiers. Curtin. ....746–52; disc. 752
(Electronics) D-C Graphical Analysis of Junction Transistor Flip-Flops. Bashkow..............1–7
(Electronics) Field Effect Transistor Applications. Huang, Marshall, White....................323–9
Electronics in Telephone Switching Systems. Joel.... ....701–09; disc. 710
(Electronics) Magnetic Regulation Transistor Power Supply. Lyons..........................643–5
(Electronics) Optimum Design of Common-Emitter Transistor Audio Amplifiers. Vallese.......391–5
(Electronics) Power Transistor Switching Circuit. Huang, Slobodzinski.......................290–6
(Electronics) Practical Applications of Electrostatic Phenomena to Particulate Matter. Ralston..155–9
(Electronics) Styroflex Aluminum-Sheathed Air-Dielectric Cable. Merrell, McKean, Arbuthnott...... ....669–74; disc. 674
(Electronics) Synchronization Accuracy Obtainable with Multiplier Phototubes. Levi.........603–06
(Electronics) Transistor Analogue Computing Amplifiers for Flight Simulators. Weyrick.......338–42
(Electronics) Transistor Voltage Regulator. Spencer, Gray................................15–17
Electrostatic Phenomena to Particulate Matter, Practical Applications of. Ralston..............155–9
Emitter Transistor Servo Preamplifiers, A Comparison of Three Common. DeSautels............17–25
Engineering Aspects of the Machine Translation of Language, Some of the. Wall....580–4; disc. 585
Engineer's Needs in Information Theory, The Communication. Rea..........................805–08
Equations on the 704 EDPM, Numerical Integration of Differential. Anderson, Johnson..........569–74
Equipment Descriptions of Type-P1 Carrier System, Circuit and. Eberhart, Hallenbeck, Perkins..... 195–204
Equipment for Radio-Interference Reduction, Design of Electronic. Albin, Sachs..................352–7
Equipment for the North Atlantic Link, Power-Feed. Meszaros, Spencer.........................890–8
Equivalent Network, Design of Wide-Band R-F Transformers Utilizing a Synthesized. Kajihara..802–05
Error Matrices, Analogue Computer Synthesis and. Honnell, Horn..........................26–32
Evaluation of Electrical Components, The Use of Computing Machines in the Statistical. Andrus.501–05
Evaluation of Electromagnets, The Analytical Design and. Kelly, Wallace......................675–80
Experience with the A2A Video System, Field. Gast.. 710–16

F

Facsimile Telegraph Transmitter, A Flat-Bed. Buckingham..........................131–3
Ferroelectric Memory Capacitors, Signals from Switched. Pulvari, McDuffie.......................681–5
Ferroelectric Resonant Trigger Pair, An Analysis and Design of the. Gremer.....................404–07
Ferroresonant Circuit with Coil Hysteresis, Subharmonic Response of the. Brenner................450–6
Ferroresonant Trigger Pair: Analysis and Design, The. Spitzer..........................407–16
Field Effect Transistor Applications. Huang, Marshall, White..........................323–9
Field Experience with the A2A Video System. Gast.. 710–16
Field Strength and Current Density in Electrical Precipitation, A New Technique for the Measurement of Corona. Cooperman..........................64–7
Field Survey of Transformer Oil Quality, A. Degnan, Shimanski..................575–9; disc. 579
Fields in Materials with Rectangular Hysteresis Loops, Superposed Magnetic. Wakeman, Beck.....562–9
Finishes, Arc-Resistant Molding Materials and. Sterling..........................776–80
Finite Pulse Width: Open Loop, Analysis of Linear Sampled-Data Systems with. Farmanfarma..... 808–19
First Look at Random Noise, A. Rice..........128–31
Five-Digit Plugboard System for Card-Programmed Calculator, A New. Williams, Noonan......67–71
Flat-Bed Facsimile Telegraph Transmitter, A. Buckingham..........................131–3
Flicker, An Instrument to Record Voltage. Kraybill, Jensen..................527–30; disc. 530
Flight Simulators, Transistor Analogue Computing Amplifiers for. Weyrick..................338–42
Flip-Flops, D-C Graphical Analysis of Junction Transistor. Bashkow..........................1–7
Flux Reversal in Magnetic Amplifier Cores. Friedlaender..................268–76; disc. 277
Formulas in Inch Units of Length, Electrical. Williamson..........................61–4
Four-Channel Open-Wire Carrier System Employing Transistors, A. Caruthers, Chaskin, Krengel.... 314–22
4,000-Mile Transmission System, Application of Statistical Techniques to Electron Tubes for Use in a. Van Haste..........................50–4
Four-Wire Switching of Long-Distance Telephone Circuits, Recent Developments in. Molnar......... ....722–32; disc. 732
Fourier Series Coefficients, Graphical Determination of the. Middendorf..........................478–82
Frequency Multipliers and Their Rating, Magnetic. McMurray:
  Part I. Frequency Triplers..................384–90
Frequency Networks, Telephone Carrier. Wallenstein 691–7
Full-Wave Bridge Magnetic Amplifiers with Inductive Loads. Woodson..........................7–9
Function Generation, Optimum Linear-Segment. Hamer..........................518–20
Fused Silicon Rectifier, The. Henkels.........733–46

G

Galvanometer Efficiency as a Design Parameter. Harris. ....176–80; disc. 180
Generation of Probability Distributions for Operations Research, Analogue Computer. Diamantides.... 86–91
Generation, Optimum Linear-Segment Function. Hamer..........................518–20
Generator Practical Design Limitations, Magnetic-Pulse-. Thompson, Trautwein, Ingersoll...786–92
Germanium and Silicon Rectifiers, Rating and Application of. Gutzwiller..........................753–7
Germanium Power Rectifier, Direct Water-Cooled. Wahl..........................832–41
Germanium Rectifiers for Industrial Applications. Burton..........................41–4
Graphical Analysis of Junction Transistor Flip-Flops, D-C. Bashkow..........................1–7
Graphical Determination of the Fourier Series Coefficients. Middendorf..........................478–82

H

Half-Duplex Teletypewriter Lines, Automatic Dispatch System for. Maxwell, Farrar..............441–5
Half-Wave Bridge Magnetic Amplifier, Alternating Current Control of the. Bernstein, Schmitz..433–8
Head Design, Magnetic Data Recording Theory:. Hoagland..........................506–12
Heat as a Means of Air Purification. Beckett, Clifton.. 373–7
High-Speed Magnetic Servo Amplifiers with 2-Phase Motor Load, Measurements on. Geyger....242–8
High-Temperature Selenium Rectifiers—A Survey of Manufacturers' Data. Shilliday.248–51; disc. 251
Hysteresis Loop Core Material in Series with Capacitors, Analysis of Instability and Response of Reactors with Rectangular. Salihi......296–305; disc. 305
Hysteresis Loops in Dielectric Amplifiers. Wingrove, Depian, Shevel..................283–8; disc. 288
Hysteresis Loops, Superposed Magnetic Fields in Materials with Rectangular. Wakeman, Beck...562–9
Hysteresis, Subharmonic Response of the Ferroresonant Circuit with Coil. Brenner..................450–6

I

IBM Type 650, The Analysis of Sampled Data Servomechanisms Performed on the. Tostanoski.446–50
Impedance of Thin-Wire Loop Antennas. Storer.... 606–18
Impedance Operational Characteristics of Toroidal Cores Used in Magnetic Amplifiers, Low-. Boyajian..........................307–13
Impedance Tensor of the General Machine, The. Yu.. 181–7
Impedances of Nonuniform Lines by a Direct Method, Computation of the. Pipes..................551–4
Inch Units of Length, Electrical Formulas in. Williamson..........................61–4
Inductions, The Use of an A-C Bridge to Measure Core Loss at High. Cooter, Harris.....252–5; disc. 256
Inductive Loads, Full-Wave Bridge Magnetic Amplifiers with. Woodson..........................7–9
Industrial Applications, Germanium Rectifiers for Burton..........................41–4
Influence of Noise on Telephone Signaling Circuit Performance. Weber..........................636–42
Information Theory, The Communication Engineer's Needs in. Rea..........................805–08
Instability and Response of Reactors with Rectangular Hysteresis Loop Core Material in Series with Capacitors, Analysis of. Salihi....296–305; disc. 305
Instrument to Record Voltage Flicker, An. Kraybill, Jensen..................527–30; disc. 530
(Instruments) Galvanometer Efficiency as a Design Parameter. Harris............176–80; disc. 180
Insulations, Nontracking Organic. Norman, Pfuntner, Kessel..........................257–61
Integrated Data Processing, Control System for. Easterlin..........................421–7
Integrated s-Plane Synthesis Using 2-Way Root Locus. Zaborszky..........................797–801
Integration of Differential Equations on the 704 EDPM, Numerical. Anderson, Johnson............569–74
Intensities, Magnetic Effects of Compressional Stress at Low Field. Fischell............148–51; disc. 151
Interference Reduction, Design of Electronic Equipment for Radio-. Albin, Sachs..................352–7
Inverse-Time Underfrequency Relay. Hoft......625–9
Investigations Concerning Polarization in Barium Titanate Ceramics. Marks, Waidelich, Monson... 469–77
Iron-Cored Oscillatory Circuits, A New Method for Treating Nonlinear Problems with Applications to. Gohar:
  Part I. Subharmonics Do Not Exist.........771–6

J

Junction Transistors to Carrier-Frequency Computing Amplifiers, Application of. Curtin............ ....746–52; disc. 752

K

Kelvin Bridge for the Audio-Frequency Range, An A-C. Dunfee..................123–7; disc. 127

L

Ladders and ar. Ideal Transformer, Synthesis of Transfer Functions with Poles Restricted to the Negative Real Axis into Two Parallel R-C. Malti, Sun.... ....165–71; disc. 171
Language, Some of the Engineering Aspects of the Machine Translation of. Wall.......580–4; disc. 585
Languages by Machine, The Technical Feasibility of Translating. Yngve..........................792–7
Length, Electrical Formulas in Inch Units of. Williamson..........................61–4
Lighting from Black and White to Color Television, Conversion of Studio. Wygant..............75–8
Lightning Protection on the Stevens Point–Wisconsin Rapids Intercity Telephone Cable. Dowling.... 697–701
Line Voltage Regulator Having Magnetic Amplifier Control, A. Anderson..........................108–12
Linear Elements, Representation of Nonlinear Characteristics by. Abou-Hussein..................84–6
Linear Programming on an Electronic Analogue Computer. Pyne..................139–42; disc. 142
Linear Sampled-Data Systems with Finite Pulse Width: Open Loop, Analysis of. Farmanfarma....808–19
Linear-Segment Function Generation, Optimum. Hamer..........................518–20
Lines by a Direct Method, Computation of the Impedances of Nonuniform. Pipes..................551–4
Load, Standing-Wave Ratio of Inaccessible. Polk.9–11
Loads, Full-Wave Bridge Magnetic Amplifiers with Inductive. Woodson..........................7–9
Loci for Control System Design, Some Mathematical Properties of Root. Reza..................103–08
Logic Circuit Control System Design Considerations, Magnetic. Van Nice..................595–600
Long-Distance Telephone Circuits, Recent Developments in 4-Wire Switching of. Molnar.722–32; disc. 732
Loop Antennas, Impedance of Thin-Wire. Storer.... 606–18
Loops, Superposed Magnetic Fields in Materials with Rectangular Hysteresis. Wakeman, Loops...562–9
Low Field Intensities, Magnetic Effects of Compressional Stress at. Fischell............148–51; disc. 151
Low-Impedance Operational Characteristics of Toroidal Cores Used in Magnetic Amplifiers. Boyajian... 307–13

M

Machine, A Data Transmission. Doty, Tate...600–03
Machine Data Over Standard Telegraph Channels, Transmission of Business. Bramhall............ ....416–20; disc. 420
Machine, The Impedance Tensor of the General. Yu. 181–7
Machine, The Technical Feasibility of Translating Languages by. Yngve..........................792–7
Machine Translation of Language, Some of the Engineering Aspects of the. Wall.....580–4; disc. 585
Machines in the Statistical Evaluation of Electrical Components, The Use of Computing. Andrus. 501–05
Magnetic Amplifier, A Signal-Discriminating. Lynn, Ringelman..........................97–102
Magnetic Amplifier, Alternating Current Control of the Half-Wave Bridge. Bernstein, Schmitz...433–8

Magnetic Amplifier Control, A Line-Voltage Regulator Having. Anderson...108-12
Magnetic Amplifier Control of Switching Transistors. Collins...585-9
Magnetic Amplifier Cores, Flux Reversal in. Friedlaender...268-76; disc. 277
Magnetic Amplifier, Dynamic Behavior of a 3-Phase. Ellert...554-60; disc. 560
Magnetic Amplifier Switching Matrix, A. Katz.236-41
Magnetic Amplifier, The Operation of the Self-Balancing. Krall, Hooper...79-84
Magnetic Amplifier, The Series. Barker:
  Parts I and II...819-30; disc. 830
(Magnetic Amplifiers) Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop Core Material in Series with Capacitors. Salihi...296-305; disc. 305
Magnetic Amplifiers, Diode-Shunting in. Lowrance, Dolan...619-24
(Magnetic Amplifiers) Hysteresis Loops in Dielectric Amplifiers. Wingrove, Depian, Shevel...283-8; disc. 288
(Magnetic Amplifiers) Inverse-Time Under frequency Relay. Hoft...625-9
Magnetic Amplifiers, Low-Impedance Operational Characteristics of Toroidal Cores Used in. Boyajian...307-13
(Magnetic Amplifiers) The Pulse Stretch Coupling Circuit. Patton...377-9
(Magnetic Amplifiers) The Simple Reactor Circuit: Its Operation and Mode Transition. Ringelman, Fenaroli...660-8; disc. 668
(Magnetic Amplifiers) The Triductor. Biringer...590-3; disc. 593
Magnetic Amplifiers with Inductive Loads, Full-Wave Bridge. Woodson...7-9
Magnetic Core Material, Dynamax, a New Crystal and Domain-Oriented. Howe...548-50; disc. 551
Magnetic Cores, Transcendental Function Analogue Computation with. Schaefer, Van Allen...160-4; disc. 164
Magnetic-Coupled Oscillator, A Single-Transistor. Chen, Schiewe...396-9; disc. 400
Magnetic Data Recording Theory: Head Design. Hoagland...506-12
Magnetic Effects of Compressional Stress at Low Field Intensities. Fischell...148-51; disc. 151
Magnetic Field, Properties of a D-C Arc in a. Winsor, Lee...143-8; disc. 148
Magnetic Fields in Materials with Rectangular Hysteresis Loops, Superposed. Wakeman, Beck...562-9
Magnetic Frequency Multipliers and Their Rating. McMurray:
  Part I. Frequency Triplers...384-90
Magnetic Logic Circuit Control System Design Considerations. Van Nice...595-600
Magnetic-Pulse-Generator Practical Design Limitations Thompson, Trautwein, Ingersoll...786-92
Magnetic Regulation Transistor Power Supply. Lyons...643-5
Magnetic Servo Amplifiers with 2-Phase Motor Load, Measurements on High-Speed. Geyger...242-8
Manufacture of Wire Spring Relays for Communication Switching Systems. Rice...513-18
Manufacturers' Data, High-Temperature Selenium Rectifiers—A Survey of. Shilliday..248-51; disc. 251
Manufacturing Facilities for Reed Switches, The Development of Automatic. Hosford...496-500
Material, Dynamax, a New Crystal and Domain-Oriented Magnetic Core. Howe..548-50; disc. 551
Materials and Finishes, Arc-Resistant Molding. Sterling...776-80
Materials with Rectangular Hysteresis Loops, Superposed Magnetic Fields in. Wakeman, Beck..562-9
Mathematical Properties of Root Loci for Control System Design, Some. Reza...103-08
Matrices, Analogue Computer Synthesis and Error. Honnell, Horn...26-32
Matrix, A Magnetic Amplifier Switching. Katz.236-41
Matter, Practical Applications of Electrostatic Phenomena to Particulate. Ralston...155-9
Measurement of Corona Field Strength and Current Density in Electrical Precipitation, A New Technique for the. Cooperman...64-7
(Measurements) An A-C Kelvin Bridge for the Audio Frequency Range. Dunfee...123-7; disc. 127
Measurements on High-Speed Magnetic Servo Amplifiers with 2-Phase Motor Load. Geyger...242-8
(Measurements) Two New Total Radiation Pyrometers. Derganc, Howell...360-7
Measuring 1,800,000-Kw Demand. Morris, Thomas...95-7
Measuring Set with Sync Pulses, A Video Visual. Sproul...427-32
Measuring Spot Size of Cathode-Ray Tube, Method of. Kuhn, Levine...357-9; disc. 359
(Megacycles) Propagation Test on 955.5 Mc, 1,965 Mc, and 6,700 Mc. Mathwich, Nuttall, Pitman, Randolph...685-91
Memory Capacitors, Signals from Switched Ferroelectric. Pulvari, McDuffie...681-5
(Metallic Rectifiers) Rating and Application of Germanium and Silicon Rectifiers. Gutzwiller..753-7
Method for Treating Nonlinear Problems with Applications to Iron-Cored Oscillatory Circuits, A New. Gohar:
  Part I. Subharmonics Do Not Exist...771-6
Method of Measuring Spot Size of Cathode-Ray Tube. Kuhn, Levine...357-9; disc. 359
Method of Nonlinear Network Analysis, A Phasor. Neal...630-6; disc. 636
Microwave Facilities with Built-In Reliability. Kuck...438-41
Microwave for Telephone Companies. Fisher...538-42
Minneapolis Tests, Co-ordinated Protection for Open-Wire Joint Use—. Albrecht, Dietz, Christoferson, Slothower...217-22; disc. 223
Mode of Operation, Transistor Power Amplifiers with Switched. Milnes...368-72
Mode Transition, The Simple Reactor Circuit; Its Operation and. Ringelman, Fenaroli...660-8; disc. 668
Molding Materials and Finishes, Arc-Resistant. Sterling...776-80
Motor Load, Measurements on High-Speed Magnetic Servo Amplifiers with 2-Phase. Geyger...242-8
Multiplex Telegraph Set, A Transistorized Time-Division. Biggam...780-5
Multiplier Phototubes, Synchronization Accuracy Obtainable with. Levi...603-06
Multipliers and Their Rating, Magnetic Frequency. McMurray:
  Part I. Frequency Triplers...384-90
Multiunit Neutralizing Transformers. Palmquist...717-21

N

Network Analysis, A Phasor Method of Nonlinear. Neal...630-6; disc. 636
Network Analysis with a Digital Computation System, Automatic. Wong, Kochen...172-5; disc. 176
Network, Design of Wide Band R-F Transformers Utilizing a Synthesized Equivalent. Kajihara...802-05
Networks, Telephone Carrier Frequency. Wallenstein...691-7
Neutralizing Transformers, Multiunit. Palmquist...717-21
Newfoundland–Nova Scotia Link, Power-Feed System for the. Thomas, Kelly...938-43
Newfoundland–Nova Scotia Link, Repeater Design for the. Brockbank, Walker, Welsby...927-38
Newfoundland–Nova Scotia Link, System Design for the. Halsey, Bampton...917-27
Noise, A First Look at Random. Rice...128-31
Noise on Telephone Signaling Circuit Performance, Influence of. Weber...636-42
Noise Reduction in Communication Channels, Principles of. Schwartz...44-50
Nonlinear Characteristics by Linear Elements, Representation of. Abou-Hussein...84-6
Nonlinear Network Analysis, A Phasor Method of. Neal...630-6; disc. 636
Nonlinear Problems with Applications to Iron-Cored Oscillatory Circuits, A New Method for Treating. Gohar:
  Part I. Subharmonics Do Not Exist...771-6
Nontracking Organic Insulations. Norman, Pfuntner, Kessel...257-61
Nonuniform Lines by a Direct Method, Computation of the Impedances of. Pipes...551-4
North Atlantic Link, Power-Feed Equipment for the. Meszaros, Spencer...890-8
North Atlantic Link, Repeater Design for the. Gleichmann, Lince, Wooley, Braga...865-77
North Atlantic Link, Repeater Production for the. Lamb, Heffner...877-89
North Atlantic Link, System Design for the. Lewis, Tucker, Lovell, Fraser...850-64
Note on the Approximation Problem. Balabanian...72-5
Nova Scotia Link, Power-Feed System for the Newfoundland–. Thomas, Kelly...938-43
Nova Scotia Link, Repeater Design for the Newfoundland–. Brockbank, Walker, Welsby...927-38
Nova Scotia Link, System Design for the Newfoundland–. Halsey, Bampton...917-27
Numerical Integration of Differential Equations on the 704 EDPM. Anderson, Johnson...569-74

O

Objectives and General Description of the Type-P1 Carrier System. Boyd...188-91
Oil Quality, A Field Survey of Transformer. Degnan, Shimanski...575-9; disc. 579
1,800,000-Kw Demand, Measuring. Morris, Thomas...95-7
Ontario Tests, Co-ordinated Protection for Open-Wire Joint Use—. Ellis, Phelps, Roach, Treen...223-34; disc. 235
Open-Wire Carrier System Employing Transistors, A New 4-Channel. Caruthers, Chaskin, Krengel...314-22
Open-Wire Joint Use—Minneapolis Tests, Co-ordinated Protection for. Albrecht, Dietz, Christoferson, Slothower...217-22; disc. 223
Open-Wire Joint Use—Ontario Tests, Co-ordinated Protection for. Ellis, Phelps, Roach, Treen...223-34; disc. 235
Open-Wire Joint Use—Present Trends, Co-ordinated Protection for. Bullard, Weppler...215-16
Operation of the Self-Balancing Magnetic Amplifier, The. Krall, Hooper...79-84
Operational Characteristics of Toroidal Cores Used in Magnetic Amplifiers, Low-Impedance. Boyajian...307-13
Operations Research, Analogue Computer Generation of Probability Distributions for. Diamantides...86-91
Optimum Design of Common-Emitter Transistor Audio Amplifiers. Vallese...391-5
Optimum Linear-Segment Function Generation. Hamer...518-20
Oregon Radio System for the Bonneville Power Administration, Southwest. Marihart, Wylie..531-8
Organic Insulations, Nontracking. Norman, Pfuntner, Kessel...257-61
Oscillator, A Single-Transistor Magnetic-Coupled. Chen, Schiewe...396-9; disc. 400
Oscillatory Circuits, A New Method for Treating Nonlinear Problems with Applications to Iron-Cored. Gohar:
  Part I. Subharmonics Do Not Exist...771-6

P

P1 Carrier System, Circuit, Circuit and Equipment Descriptions of Type-. Eberhart, Hallenbeck, Perkins...195-204
P1 Carrier System, Objectives and General Description of the Type. Boyd...188-91
P1 Carrier System, Power Supplies for the Type-. Smith...191-4; disc. 194
P1 Carrier System to Rural Telephone Lines, Application of the Type-. Howard...205-13; disc. 214
Parallel R-C Ladders and an Ideal Transformer, Synthesis of Transfer Functions with Poles Restricted to the Negative Real Axis into Two. Malti, Sun...165-71; disc. 171
Parameter, Galvanometer Efficiency as a Design. Harris...176-80; disc. 180
Particulate Matter, Practical Applications of Electrostatic Phenomena to. Ralston...155-9
Performance Requirements, Color-Television System. Kennedy...653-9
Phasor Method of Nonlinear Network Analysis, A. Neal...630-6; disc. 636
Phenomena, Physical Response Requirements of Pressure Transducers for the Reproduction of Physiological. Wood...32-40
Phenomena to Particulate Matter, Practical Applications of Electrostatic. Ralston...155-9
Phototubes, Synchronization Accuracy Obtainable with Multiplier. Levi...603-06
Physical Response Requirements of Pressure Transducers for the Reproduction of Physiological Phenomena. Wood...32-40
Physiological Phenomena, Physical Response Requirements of Pressure Transducers for the Reproduction of. Wood...32-40
Plate, Accurate Determination of the Capacitance of a Thin Rectangular. Reitan, Higgins.761-6; disc. 766
Plugboard System for Card-Programmed Calculator A New Five-Digit. Williams, Noonan...67-71
Polarization in Barium Titanate Ceramics, Investigations Concerning. Marks, Waidelich, Monson..469-77
Poles Restricted to the Negative Real Axis into Two Parallel R-C Ladders and an Ideal Transformer, Synthesis of Transfer Functions with. Malti, Sun...165-71; disc. 171
Power Amplifiers with Switched Mode of Operation, Transistor. Milnes...368-72
Power-Feed Equipment for the North Atlantic Link. Meszaros, Spencer...890-8
Power-Feed System for the Newfoundland–Nova Scotia Link. Thomas, Kelly...938-43
Power Supplies for the Type-P1 Carrier System. Smith...191-4; disc. 194
Power Supply, Magnetic Regulation Transistor. Lyons...643-5
Power Transistor Switching Circuit. Huang, Slobodzinski...290-6
Practical Applications of Electrostatic Phenomena to Particulate Matter. Ralston...155-9
Preamplifiers, A Comparison of Three Common Emitter Transistor Servo. DeSautels...17-25
Precipitation, A New Technique for the Measurement of Corona Field Strength and Current Density in Electrical. Cooperman...64-7
Pressure Transducers for the Reproduction of Physiological Phenomena, Physical Response Requirements of. Wood...32-40

Principles of Noise Reduction in Communication Channels. Schwartz........................44–50
Printing Telegraph Switching and Integrated Data Processing, Simplified. Booth, Klich.......11–14
Probability Distributions for Operations Research Analogue Computer Generation of. Diamantides. 86–91
Problem, Note on the Approximation. Balabanian.... 72–5
Problems with Applications to Iron-Cored Oscillatory Circuits, A New Method for Treating Nonlinear. Gohar:
  Part I. Subharmonics Do Not Exist.........771–6
Processing, Control System for Integrated Data. Easterlin..421–7
Programming on an Electronic Analogue Computer, Linear. Pyne..................139–42; disc. 142
Propagation Test on 955.5 Mc, 1,965 Mc, and 6,700 Mc. Mathwich, Nuttall, Pitman, Randolph.....685–91
Properties of a D-C Arc in a Magnetic Field. Winsor, Lee..........................143–8; disc. 148
Protection for Open-Wire Joint Use—Minneapolis Tests, Co-ordinated. Albrecht, Dietz, Christoferson, Slothower................217–22; disc. 223
Protection for Open-Wire Joint Use—Ontario Tests Co-ordinated. Ellis, Phelps, Roach, Treen....... ....223–34; disc. 235
Protection for Open-Wire Joint Use—Present Trends, Co-ordinated. Bullard, Weppler..........215–16
Protection on the Stevens Point–Wisconsin Rapids Intercity Telephone Cable, Lightning. Dowling.. 697–701
Protection Switching for TD-2 Radio System, Automatic. Klemm..........................520–7.
Pulse-Generator Practical Design Limitations, Magnetic-. Thompson, Trautwein, Ingersoll...786–92
Pulse Stretch Coupling Circuit, The. Patton...377–9
Pulse Width: Open Loop, Analysis of Linear Sampled-Data Systems with Finite. Farmanfarma...808–19
Pulses, A Video Visual Measuring Set with Sync. Sproul.................................427–32
Purification, Heat as a Means of Air. Beckett, Clifton. 373–7
Pyrometers, Two New Total Radiation. Derganc, Howell...................................360–7

R

R-F Transformers Utilizing a Synthesized Equivalent Network, Design of Wide-Band. Kajihara..802–05
Radiation Pyrometers, Two New Total. Derganc, Howell...................................360–7
(Radio Communication) Design of Wide-Band R-F Transformers Utilizing a Synthesized Equivalent Network. Kajihara..........................802–05
(Radio Communication) Microwave Facilities with Built-In Reliability. Kuck................438–41
(Radio Communication) Microwave for Telephone Companies. Fisher.........................538–42
(Radio Communication) Propagation Test on 955.5 Mc, 1,965 Mc, and 6,700 Mc. Mathwich, Nuttall, Pitman, Randolph..........................685–91
(Radio Communication) Standing-Wave Ratio of Inaccessible Load. Polk.......................9–11
Radio-Interference Reduction, Design of Electronic Equipment for. Albin, Sachs..............352–7
Radio System, Automatic Protection Switching for TD-2. Klemm...................................520–7
Radio System for the Bonneville Power Administration, Southwest Oregon. Marihart, Wylie......531–8
Radio System for Toll Telephone Service, A 72-Channel. Harp, Kebby...................113–19; disc. 119
Random Noise, A First Look at. Rice..........128–31
Rating and Application of Germanium and Silicon Rectifiers. Gutzwiller..........................753–7
Rating, Magnetic Frequency Multipliers and Their. McMurray:
  Part I. Frequency Triplers...............384–90
Rating of Telephone Sets, Transmission. Johnson, Eppes..........................646–52; disc. 652
Ratio of Inaccessible Load, Standing-Wave. Polk.9–11
Reactance Calculations with Digital Computers, Transformer. Rabins...............261–7; disc. 267
Reactor Circuit; Its Operation and Mode Transition, The Simple. Ringelman, Fenaroli.660–8; disc. 668
Reactors with Rectangular Hysteresis Loop Core Material in Series with Capacitors, Analysis of Instability and Response of. Salihi..296–305; disc. 305
Recent Developments in 4-Wire Switching of Long Distance Telephone Circuits. Molnar........... ....722–32; disc. 732
Recent New Features for the No. 5 Crossbar Switching System. Dehn, Hersey.................457–66
Record Voltage Flicker, An Instrument to. Kraybill, Jensen......................527–30; disc. 530
Recording Theory: Head Design, Magnetic Data. Hoagland.................................506–12
Rectangular Hysteresis Loops, Superposed Magnetic Fields in Materials with. Wakeman, Beck..562–9
Rectangular Plate, Accurate Determination of the Capacitance of a Thin. Reitan, Higgins......... ....761–6; disc. 766
Rectifier, Direct Water-Cooled Germanium Power. Wahl.....................................832–41
Rectifier, The Fused Silicon. Henkels.........733–46
Rectifiers—A Survey of Manufacturers' Data, High-Temperature Selenium. Shilliday.............. ....248–51; disc. 251
Rectifiers for Industrial Applications, Germanium. Burton.....................................41–4
Rectifiers, Rating and Application of Germanium and Silicon. Gutzwiller.........................753–7
Reed Switches, The Development of Automatic Manufacturing Facilities for. Hosford.........496–500
Regulator Having Magnetic Amplifier Control, A Line Voltage. Anderson....................108–12
Regulator, Transistor Voltage. Spencer, Gray..15–17
Relay, Inverse-Time Underfrequency. Hoft...625–9
Relays, Capability of Sealed Contact. Hovgaard.466–8
Relays, Design Features of Bell System Wire Spring. Knapp.....................................482–6
Relays for Communication Switching Systems, Manufacture of Wire Spring. Rice.............513–18
Reliability, Microwave Facilities with Built-In. Kuck. 438–41
Repeater Design for the Newfoundland–Nova Scotia Link. Brockbank, Walker, Welsby.......927–38
Repeater Design of the North Atlantic Link. Gleichmann, Lince, Wooley, Braga..............865–77
Repeater Production for the North Atlantic Link. Lamb, Heffner.............................877–89
Representation of Nonlinear Characteristics by Linear Elements. Abou-Hussein..................84–6
Research, Analogue Computer Generation of Probability Distributions for Operations. Diamantides. 86–9
Research Project, The Wisconsin Integrally Synchronized Computer—A University. Asmuth, Davidson, Miller, Noble, Scidmore..............330–8
Resonant Trigger Pair, An Analysis and Design of the Ferroelectric. Gremer..................404–07
Response of the Ferroresonant Circuit with Coil Hysteresis, Subharmonic. Brenner.............450–6
Reversal in Magnetic Amplifier Cores, Flux. Friedlaender......................268–76; disc. 277
Root Loci for Control System Design, Some Mathematical Properties of. Reza..............103–08
Root Locus, Integrated s-Plane Synthesis Using 2-way. Zaborszky...............................797–801
Route Selection and Cable Laying for the Transatlantic Cable System. Jack, Leech, Lewis........944–55

S

s-Plane Synthesis Using 2-Way Root Locus, Integrated. Zaborszky...............................797–801
Sampled Data Servomechanisms Performed on the IBM Type 650, The Analysis of. Tostanoski.... 446–50
Sampled-Data Systems with Finite Pulse Width: Open Loop, Analysis of Linear. Farmanfarma.... 808–19
Sealed Contact Relays, Capability of. Hovgaard..... 466–8
Segment Function Generation, Optimum Linear-. Hamer...................................518–20
Selenium Rectifiers—A Survey of Manufacturers' Data, High-Temperature. Shilliday............ ....248–51; disc. 251
Self-Balancing Magnetic Amplifier, The Operation of the. Krall, Hooper........................79–84
Series Magnetic Amplifier, The. Barker:
  Parts I and II.................819–30; disc. 830
Servo Amplifiers with 2-Phase Motor Load, Measurements on High-Speed Magnetic. Geyger..242–8
Servo Preamplifiers, A Comparison of Three Common Emitter Transistor. DeSautels...........17–25
Servomechanisms Performed on the IBM Type 650, The Analysis of Sampled Data. Tostanoski....446–50
72-Channel Radio System for Toll Telephone Service, A. Harp, Kebby..............113–19; disc. 119
704 EDPM, Numerical Integration of Differential Equations on the. Anderson, Johnson.....569–74
Several-Valued Combinational Switching Circuits. Lee, Chen...............................278–83
Shunting in Magnetic Amplifiers, Diode-. Lowrance, Dolan.....................................619–24
Signal-Discriminating Magnetic Amplifier, A. Lynn, Ringelman................................97–102
Signaling Circuit Performance, Influence of Noise on Telephone. Weber..........................636–42
Signals from Switched Ferroelectric Memory Capacitors. Pulvari, McDuffie........................681–5
Silicon Rectifier, The Fused. Henkels.........733–46
Silicon Rectifiers, Rating and Application of Germanium. Gutzwiller.........................753–7
Simplified Printing Telegraph Switching and Integrated Data Processing. Booth, Klich...........11–14
Simulators, Transistor Analogue Computing Amplifiers for Flight. Weyrick.........................338–42
Single-Transistor Magnetic-Coupled Oscillator, A. Chen, Schiewe.......................396–9; disc. 400
Size of Cathode-Ray Tube, Method of Measuring Spot. Kuhn, Levine...................357–9; disc. 359
(Solid State Devices) Temperature Stabilization of Transistor Amplifiers. Vallese...........379–84
(Solid State Devices) The Symmetrical Transistor as a Bilateral Switching Element. Troudale..400–03
Some Mathematical Properties of Root Loci for Control System Design. Reza.....................103–08
Some of the Engineering Aspects of the Machine Translation of Lauguage. Wall........580–4; disc. 585
Southwest Oregon Radio System for the Bonneville Power Administration. Marihart, Wylie...531–8
Spot Size of Cathode-Ray Tube, Method of Measuring. Kuhn, Levine..................357–9; disc. 359
Spring Relays, Design Features of Bell System Wire. Knapp.....................................482–6
Spring Relays for Communication Switching Systems, Manufacture of Wire. Rice.............513–18
Stabilization of Transistor Amplifiers, Temperature. Vallese....................................379–84
Standing-Wave Ratio of Inaccessible Load. Polk..9–11
Static Elimination on Bell System Teletypewriters. Bavley, Perry, Swezey......................92–4
Statistical Evaluation of Electrical Components, The Use of Computing Machines in the. Andrus.501–05
Statistical Techniques to Electron Tubes for Use in a 4,000 Mile Transmission System, Application of. Van Haste....................................50–4
Stature of the Crossbar Tandem Switching System, The Full. Mesar............................486–96
Stevens Point—Wisconsin Rapids Intercity Telephone Cable, Lightning Protection on. Dowling..697–701
Stretch Coupling Circuit, The Pulse. Patton...377–9
Studio Lighting from Black and White to Color Television, Conversion of. Wygant...........75–8
Styroflex Aluminum-Sheathed Air-Dielectric Cable. Merrell, McKean, Arbuthnott...669–74; disc. 674
Subharmonic Response of the Ferroresonant Circuit with Coil Hysteresis. Brenner.............450–6
Submarine Cable System, Cable Design and Manufacture for the Transatlantic. Lebert, Fischer, Blakeborn...................................907–17
Superposed Magnetic Fields in Materials with Rectangular Hysteresis Loops. Wakeman, Beck.562–9
Survey of Transformer Oil Quality, A Field. Degnan, Shimanski.....................575–9; disc. 579
Switched Ferroelectric Memory Capacitors, Signals from. Pulvari, McDuffie..................681–5
Switches, The Development of Automatic Manufacturing Facilities for Reed. Hosford......496–500
Switching and Integrated Data Processing, Simplified Printing Telegraph. Booth, Klich..........11–14
Switching Circuit, Power Transistor. Huang, Slobodzinski...................................290–6
Switching Circuits, Several Valued Combinational. Lee, Chen...................................278–83
Switching Element, The Symmetrical Transistor as a Bilateral. Troudale...................400–03
Switching for TD-2 Radio System, Automatic Protection. Klemm.............................520–7
Switching Matrix, A Magnetic Amplifier. Katz..36–41
Switching of Long-Distance Telephone Circuits, Recent Developments in 4-Wire. Molnar....... ....722–32; disc. 732
Switching System, Recent New Features for the No. 5 Crossbar. Dehn, Hersey.................457–66
Switching System, The Full Stature of the Crossbar Tandem. Mesar.........................486–96
Switching Systems, Electronics in Telephone. Joel..... ....701–09; disc. 710
Switching Transistors, Magnetic Amplifier Control of. Collins..................................585–9
Symmetrical Transistor as a Bilateral Switching Element, The. Troudale.........................400–03
Sync Pulses, A Video Visual Measuring Set with. Sproul....................................427–32
Synchronization Accuracy Obtainable with Multiplier Phototubes. Levi..........................603–06
Synchronized Computer—A University Research Project, The Wisconsin Integrally. Asmuth, Davidson, Miller, Noble, Scidmore.........330–8
Synthesis and Error Matrices, Analogue Computer. Honnell, Horn................................26–32
Synthesis of Transfer Functions with Poles Restricted to the Negative Real Axis into Two Parallel R-C Ladders and an Ideal Transformer. Malti, Sun.. ....165–71; disc. 171
Synthesis Using 2-Way Root Locus, Integrated s-Plane. Zaborszky...............................797–801
Synthesized Equivalent Network, Design of Wide-Band R-F Transformers Utilizing a. Kajihara...802–05
System Design for the Newfoundland–Nova Scotia Link. Halsey, Bampton..........................917–27
System Design for the North Atlantic Link. Lewis, Tucker, Lovell, Fraser.......................850–64
System for Half-Duplex Teletypewriter Lines, Automatic Dispatch. Maxwell, Farrar.............441–5

T

Tandem Switching System, The Full Stature of the Crossbar. Mesar............................486–96
TD-2 Radio System, Automatic Protection Switching for. Klemm..................................520–7

Technical Feasibility of Translating Languages by Machine, The. Yngve....................792–7
Technique for the Measurement of Corona Field Strength and Current Density in Electrical Precipitation, A New. Cooperman..................64–7
(Telegraph) Automatic Dispatch System for Half-Duplex Teletypewriter Lines. Maxwell, Farrar.. 441–5
Telegraph Channels, Transmission of Business Machine Data Over Standard. Bramhall..416–20; disc. 420
(Telegraph) Delay Distortion Correction. Cannon... 55–61
Telegraph Set, A Transistorized Time-Division Multiplex. Biggam..............................780–5
(Telegraph) Static Elimination on Bell System Teletypewriters. Bavley, Perry, Swezey..........92–4
Telegraph Switching and Integrated Data Processing, Simplified Printing. Booth, Klich..........11–14
(Telegraph Systems) A Data Transmission Machine. Doty, Tate.................................600–03
(Telegraph Systems) Control System for Integrated Data Processing. Easterlin..................421–7
Telegraph Transmitter, A Flat-Bed Facsimile. Buckingham.....................................131–3
Telephone Cable, Lightning Protection on the Stevens Point—Wisconsin Rapids Intercity. Dowling.... 697–701
Telephone Cable System—Planning and Over-all Performance, Transatlantic. Mottram, Halsey, Emling, Griffith..............................842–50
Telephone Carrier Frequency Networks. Wallenstein. 691–7
Telephone Circuits, Recent Developments in 4-Wire Switching of Long-Distance. Molnar........... ....722–32; disc. 732
Telephone Companies, Microwave for. Fisher..538–42
Telephone Lines, Application of the Type-P1 Carrier System to Rural. Howard.....205–13; disc. 214
Telephone Service, A 72-Channel Radio System for Toll. Harp, Kebby............113–19; disc. 119
Telephone Sets, Transmission Rating of. Johnson, Eppes........................646–52; disc. 652
Telephone Signaling Circuit Performance, Influence of Noise on. Weber..........................636–42
Telephone Switching Systems, Electronics in. Joel.... ....701–09; disc. 710
Teletypewriter Lines, Automatic Dispatch System for Half-Duplex. Maxwell, Farrar..............441–5
Teletypewriters, Static Elimination on Bell System. Bavley, Perry, Swezey.................... 92–4
(Television) A Video Visual Measuring Set with Sync Pulses. Sproul...........................427–32
Television, Conversion of Studio Lighting from Black and White to Color. Wygant...............75–8
(Television) Field Experience with the A2A Video System. Gast...............................710–16
Television System Performance Requirements, Color-. Kennedy...............................653–9
Temperature Stabilization of Transistor Amplifiers. Vallese.................................379–84
Tensor of the General Machine, The Impedance. Yu. 181–7
Test on 955.5 Mc, 1,965 Mc, and 6,700 Mc, Propagation. Mathwich, Nuttall, Pitman, Randolph.... 685–91
Theory: Head Design, Magnetic Data Recording. Hoagland.....................................506–12
Theory, The Communication Engineer's Needs in Information. Rea...........................805–08
Thin Rectangular Plate, Accurate Determination of the Capacitance of a. Reitan, Higgins..761–6; disc. 766
Thin-Wire Loop Antennas, Impedance of. Storer.... 606–18
Three Common Emitter Transistor Servo Preamplifiers, A Comparison of. DeSautels..............17–25
Three-Phase Magnetic Amplifier, Dynamic Behavior of a. Ellert..................554–60; disc. 560
Time-Division Multiplex Telegraph Set, A Transistorized. Biggam..............................780–5
Time Underfrequency Relay, Inverse-. Hoft...625–9
Titanate Ceramics, Investigations Concerning Polarization in Barium. Marks, Waidelich, Monson...... 469–77
Toll Telephone Service, A 72-Channel Radio System for. Harp, Kebby..............113–19; disc. 119
Toroidal Cores Used in Magnetic Amplifiers, Low-Impedance Operational Characteristics of. Boyajian.......................................307–13
Traffic Considerations in Community Dial Offices. Lanigan, McDonough...........543–7; disc. 547
Transatlantic Cable System, Electron Tubes for the. McNally, Metson, Veazie, Holmes........898–907
Transatlantic Cable System, Route Selection and Cable Laying for the. Jack, Leech, Lewis......944–55
Transatlantic Communications: A Historical Résumé. Kelly, Radley..............................841–2
(Transatlantic) Power-Feed Equipment for the North Atlantic Link. Meszaros, Spencer........890–8
(Transatlantic) Power-Feed System for the Newfoundland–Nova Scotia Link. Thomas, Kelly...938–43
(Transatlantic) Repeater Design for the Newfoundland–Nova Scotia Link. Brockbank, Walker, Welsby.. 927–38
(Transatlantic) Repeater Design for the North Atlantic Link. Gleichmann, Lince, Wooley, Braga..865–77
(Transatlantic) Repeater Production for the North Atlantic Link. Lamb, Heffner.............877–89
Transatlantic Submarine Cable System, Cable Design and Manufacture for the. Lebert, Fischer, Biskeborn..907–17
(Transatlantic) System Design for the Newfoundland–Nova Scotia Link. Halsey, Bampton.....917–27
(Transatlantic) System Design for the North Atlantic Link. Lewis, Tucker, Lovell, Fraser......850–64
Transatlantic Telephone Cable System—Planning and Over-all Performance. Mottram, Halsey, Emling, Griffith..842–50
Transcendental Function Analogue Computation with Magnetic Cores. Schaefer, Van Allen........... ....160–4; disc. 164
Transcontinental Communications in Canada, Development of. Noyes, Gaudet, Bonneville......342–52
Transducers for the Reproduction of Physiological Phenomena, Physical Response Requirements of Pressure. Wood.......................................32–40
Transfer Functions with Poles Restricted to the Negative Real Axis into Two Parallel R-C Ladders and an Ideal Transformer, Synthesis of. Malti, Sun...... ....165–71; disc. 171
Transformer Oil Quality, A Field Survey of. Degnan, Shimanski......................575–9; disc. 579
Transformer Reactance Calculations with Digital Computers. Rabins.................261–7; disc. 267
Transformer, Synthesis of Transfer Functions with Poles Restricted to the Negative Real Axis into Two Parallel R-C Ladders and an Ideal. Malti, Sun..........................165–71; disc. 171
Transformers, Multiunit Neutralizing. Palmquist..... 717–21
Transformers to Specification, Application of a Digital Computer to the Design of Power. Moore, Duboc, Zaphyr..........................134–7; disc. 137
Transformers Utilizing a Synthesized Equivalent Network, Design of Wide-Band R-F. Kajihara.802–05
Transistor Amplifiers, Temperature Stabilization of. Vallese...................................379–84
Transistor Analogue Computing Amplifiers for Flight Simulators. Weyrick......................338–42
Transistor Applications, Field Effect. Huang, Marshall, White..................................323–9
Transistor as a Bilateral Switching Element, The Symmetrical. Trousdale.....................400–03
Transistor Audio Amplifiers, Optimum Design of Common-Emitter. Vallese......................391–5
Transistor Flip-Flops, D-C Graphical Analysis of Junction. Bashkow..............................1–7
Transistor Magnetic-Coupled Oscillator, A Single-. Chen, Schiewe................396–9; disc. 400
Transistor Power Amplifiers with Switched Mode of Operation. Milnes..........................368–72
Transistor Power Supply, Magnetic Regulation. Lyons..643–5
Transistor Servo Preamplifiers, A Comparison of Three Common Emitter. DeSautels.............17–25
Transistor Switching Circuit, Power. Huang, Slobodzinski.......................................290–6
Transistor Voltage Regulator. Spencer, Gray..15–17
Transistorized Time-Division Multiplex Telegraph Set, A. Biggam...............................780–5
Transistors, A New 4-Channel Open-Wire Carrier System Employing. Caruthers, Chaskin, Krengel... 314–22
Transistors, Magnetic Amplifier Control of Switching. Collins...................................585–9
Transistors to Carrier-Frequency Computing Amplifiers, Application of Junction. Curtin.............. ....746–52; disc. 752
Translating Languages by Machine, The Technical Feasibility of. Yngve.........................792–7
Translation of Language, Some of the Engineering Aspects of the Machine. Wall..580–4; disc. 585
Transmission Machine, A Data. Doty, Tate...600–03
Transmission of Business Machine Data Over Standard Telegraph Channels. Bramhall..416–20; disc. 420
Transmission Rating of Telephone Sets. Johnson, Eppes.........................646–52; disc. 652
Transmission System, Application of Statistical Techniques to Electron Tubes for Use in a 4,000-Mile. Van Haste..50–4
Transmitter, A Flat-Bed Facsimile Telegraph. Buckingham.....................................131–3
Triductor, The. Biringer............590–3; disc. 593
Trigger Pair, An Analysis and Design of the Ferroelectric Resonant. Gremer..................404–07
Trigger Pair: Analysis and Design, The Ferroresonant. Spitzer..................................407–16
Tube, Method of Measuring Spot Size of Cathode-Ray. Kuhn, Levine..................357–9; disc. 359
Two New Total Radiation Pyrometers. Derganc, Howell..360–7
2-Phase Motor Load, Measurements on High-Speed Magnetic Servo Amplifiers with. Geyger..242–8
Type-P1 Carrier System, Circuit and Equipment Descriptions of. Eberhart, Hallenbeck, Perkins..... 195–204
Type-P1 Carrier System, Objectives and General Description of the. Boyd....................188–91
Type-P1 Carrier System, Power Supplies for the. Smith..........................191–4; disc. 194
Type-P1 Carrier System to Rural Telephone Lines, Application of the. Howard....205–13; disc. 214

U

Underfrequency Relay, Inverse-Time. Hoft....625–9
Units and Dimensions, Electrical. Young........... ....767–8; disc. 768
University Research Project, The Wisconsin Integrally Synchronized Computer—A. Asmuth, Davidson, Miller, Noble, Scidmore..........................330–8
Use of an A-C Bridge to Measure Core Loss at High Inductions, The. Cooter, Harris..252–5; disc. 256
Use of Computing Machines in the Statistical Evaluation of Electrical Components, The. Andrus...501–05

V

Video System, Field Experience with the A2A. Gast.. 710–16
Video Visual Measuring Set with Sync Pulses, A. Sproul..427–32
Visual Measuring Set with Sync Pulses, A Video. Sproul..427–32
Voltage Flicker, An Instrument to Record. Kraybill, Jensen.......................527–30; disc. 530
Voltage Regulator Having Magnetic Amplifier Control, A Line. Anderson.......................108–12
Voltage Regulator, Transistor. Spencer, Gray...15–17

W

Water-Cooled Germanium Power Rectifier, Direct. Wahl...832–41
Wide-Band R-F Transformers Utilizing a Synthesized Equivalent Network, Design of. Kajihara.802–05
(Wire Communication) A New 4-Channel Open-Wire Carrier System Employing Transistors. Caruthers, Chaskin, Krengel..............................314–22
(Wire Communication) Application of the Type-P1 Carrier System to Rural Telephone Lines. Howard ....205–13; disc. 214
(Wire Communication) Circuit and Equipment Descriptions of Type-P1 Carrier System. Eberhart, Hallenbeck, Perkins......................195–204
(Wire Communication) Co-ordinated Protection for Open-Wire Joint Use—Minneapolis Tests. Albrecht, Dietz, Christoferson, Slothower.......... ....217–22; disc. 223
(Wire Communication) Co-ordinated Protection for Open-Wire Joint Use—Ontario Tests. Ellis, Phelps, Roach, Treen..........223–34; disc. 235
(Wire Communication) Co-ordinated Protection for Open-Wire Joint Use—Present Trends. Bullard, Weppler..215–16
(Wire Communication) Development of Transcontinental Communications in Canada. Noyes, Gaudet, Bonneville.......................................342–52
(Wire Communication) Influence of Noise on Telephone Signaling Circuit Performance. Weber.... 636–42
(Wire Communication) Lightning Protection on the Stevens Point—Wisconsin Rapids Intercity Telephone Cable. Dowling......................697–70
(Wire Communication) Multiunit Neutralizing Transformers. Palmquist..........................717–21
(Wire Communication) Objectives and General Description of the Type-P1 Carrier System. Boyd.. 188–91
(Wire Communication) Power Supplies for the Type-P1 Carrier System. Smith.......191–4; disc. 194
(Wire Communication) Telephone Carrier Frequency Networks. Wallenstein.......................691–7
(Wire Communication) Transmission Rating of Telephone Sets. Johnson, Eppes....646–52; disc. 652
Wire Loop Antennas, Impedance of Thin-. Storer.... 606–18
Wire Spring Relays, Design Features of Bell System. Knapp.......................................482–6
Wire Spring Relays for Communication Switching Systems, Manufacture of. Rice.............513–18
Wisconsin Integrally Synchronized Computer—A University Research Project, The. Asmuth, Davidson, Miller, Noble, Scidmore..........................330–8
Wisconsin Rapids Intercity Telephone Cable, Lightning Protection on the Stevens Point—. Dowling..... 697–701

2. Author Index

Abetti, P. A.; S. B. Williams. Disc.......137

Abou-Hussein, M. S. M. Representation of Nonlinear Characteristics by Linear Elements. (56-121)..84–6

Albin, A. L.; H. M. Sachs. Design of Electronic Equipment for Radio-Interference Reduction. (56-257).......352–7

Albrecht, E. G.; A. E. Dietz, E. W. Christoferson, J. C. Slothower. Co-ordinated Protection for Open-Wire Joint Use—Minneapolis Tests. (56-239).......217–22; disc. 223

Anderson, F. W. A Line Voltage Regulator Having Magnetic Amplifier Control. (56-227).....108–12

Anderson, H. H.; J. R. Johnson. Numerical Integration of Differential Equations on the 704 EDPM. (56-697).......569–74

Andrus, W. E., Jr. The Use of Computing Machines in the Statistical Evaluation of Electrical Components. (56-679).......501–05

Arbuthnott, J., Jr.; E. J. Merrell, A. L. McKean. Styroflex Aluminum-Sheathed Air-Dielectric Cable. (56-803).......669–74; disc. 674

Asmuth, J. L.; C. H. Davidson, J. B. Miller, D. S. Noble, A. K. Scidmore. The Wisconsin Integrally Synchronized Computer—A University Research Project. (55-742).......330–8

Astrahan, M. M. Disc.......585

B

Balabanian, Norman. Note on the Approximation Problem. (56-128).......72–5

Bampton, J. F.; R. J. Halsey. System Design for the Newfoundland–Nova Scotia Link.......917–27

Barker, R. C. The Series Magnetic Amplifier: Parts I and II. (56-747) and (56-748).......819–30; disc. 830

Bashkow, T. R. D-C Graphical Analysis of Junction Transistor Flip-Flops. (56-24).......1–7

Bavley, Harold; D. B. Perry, B. S. Swezey. Static Elimination on Bell System Teletypewriters. (56-106).......92–4

Beck, C. J.; C. B. Wakeman. Superposed Magnetic Fields in Materials with Rectangular Hysteresis Loops. (56-717).......562–9

Beckett, J. C.; C. E. Clifton. Heat as a Means of Air Purification. (56-236).......373–7

Bernstein, T.; N. L. Schmitz. Alternating Current Control of the Half-Wave Bridge Magnetic Amplifier. (56-626).......433–8

Biggam, F. D. A Transistorized Time-Division Multiplex Telegraph Set. (56-987).......780–5

Bills, G. W. Disc.......176

Biringer, Paul P. The Triductor. (56-715).......590–3; disc. 593

Biskeborn, M. C.; A. W. Lebert, H. B. Fischer. Cable Design and Manufacture for the Transatlantic Submarine Cable System.......907–17

Bonneville, S.; J. W. Noyes, G. Gaudet. Development of Transcontinental Communications in Canada. (56-244).......342–52

Booth, J. B.; R. H. Klich. Simplified Printing Telegraph Switching and Integrated Data Processing. (56-28).......11–14

Bourne, A. C. Disc.......668

Bourne, H. C., Jr. Disc.......819

Boyajian, C. L. Low-Impedance Operational Characteristics of Toroidal Cores Used in Magnetic Amplifiers. (56-220).......307–13

Boyd, R. C. Objectives and General Description of the Type-P1 Carrier System. (56-240).......188–91

Braga, F. J.; T. F. Gleichmann, A. H. Lince, M. C. Wooley. Repeater Design for the North Atlantic Link.......865–77

Bramhall, F. D. Transmission of Business Machine Data Over Standard Telegraph Channels. (56-654).......416–20; disc. 420

Brenner, Egon. Subharmonic Response of the Ferroresonant Circuit with Coil Hysteresis. (56-686).......450–6

Brockbank, R. A.; D. C. Walker, V. G. Welsby. Repeater Design for the Newfoundland–Nova Scotia Link.......927–38

Buckingham, W. D. A Flat Bed Facsimile Telegraph Transmitter (56-108).......131–3

Bullard, W. R.; H. E. Weppler. Co-ordinated Protection for Open-Wire Joint Use—Present Trends. (56-237).......215–16

Burton, L. W. Germanium Rectifiers for Industrial Applications. (56-78).......41–4

C

Cannon, W. D. Delay Distortion Correction. (56-107).......55–61

Caruthers, R. S. Disc.......119

Caruthers, R. S.; W. S. Chaskin, H. K. Krengel. A New 4-Channel Open-Wire Carrier System Employing Transistors. (56-245).......314–22

Chambers, Carl C. Disc.......768

Chapman, F. T. Disc.......768

Chaskin, W. S.; R. S. Caruthers, H. K. Krengel. A New 4-Channel Open-Wire Carrier System Employing Transistors. (56-245).......314–22

Chen, Kan; A. J. Schiewe. A Single-Transistor Magnetic-Coupled Oscillator. (56-207).......396–9

Chen, K.; A. J. Schiewe. Disc.......164

Chen, W. H.; C. Y. Lee. Several Valued Combinational Switching Circuits. (56-162).......278–83

Christoferson, E. W.; E. G. Albrecht, A. E. Dietz, J. C. Slothower. Co-ordinated Protection for Open-Wire Joint Use—Minneapolis Tests. (56-239).......217–22; disc. 223

Clifton, C. E.; J. C. Beckett. Heat as a Means of Air Purification. (56-236).......373–7

Cogbill, B. A. Disc.......267

Collins, H. William. Magnetic Amplifier Control of Switching Transistors. (56-726).......585–9

Cone, D. I.; J. M. Moses. Disc.......214

Cooperman, P. A New Technique for the Measurement of Corona Field Strength and Current Density in Electrical Precipitation. (56-113).......64–7

Cooter, I. L.; W. P. Harris. The Use of an A-C Bridge to Measure Core Loss at High Inductions. (56-26).......252–5; disc. 256

Crocker, Charles F. Disc.......194

Curtin, W. A. Application of Junction Transistors to Carrier-Frequency Computing Amplifiers. (56-758).......746–52; disc. 752

Cybulski, John. Disc.......151, 278

D

Davidson, C. H.; J. L. Asmuth, J. B. Miller, D. S. Noble, A. K. Scidmore. The Wisconsin Integrally Synchronised Computer—A University Research Project. (56-742).......330–8

Degnan, W. J.; E. J. Shimanski. A Field Survey of Transformer Oil Quality. (56-805).......575–9; disc. 579

Dehn, J. W.; R. E. Hersey. Recent New Features for the No. 5 Crossbar Switching System. (56-667).......457–66

Depian, Louis; E. Wingrove, W. L. Shevel. Hysteresis Loops in Dielectric Amplifiers. (56-206).......283–8; disc. 289

Derganc, W.; S. N. Howell. Two New Total Radiation Pyrometers. (56-98).......360–7

DeSautels, A. N. A Comparison of Three Common Emitter Transistor Servo Preamplifiers. (56-37).......17–25

Diamantides, N. D. Analogue Computer Generation of Probability Distributions for Operations Research. (56-148).......86–91

Dieterly, D. C. Disc.......256

Dietz, A. E.; E. G. Albrecht, E. W. Christoferson, J. C. Slothower. Co-ordinated Protection for Open-Wire Joint Use—Minneapolis Tests. (56-239).......217–22; disc. 223

Dolan, J. D.; J. L. Lowrance. Diode-Shunting in Magnetic Amplifiers. (56-749).......619–24

Doty, C. R.; L. A. Tate. A Data Transmission Machine. (56-985).......600–03

Dowling, R. C. Lightning Protection on the Stevens Point–Wisconsin Rapids Intercity Telephone Cable. (56-971).......697–701

Duboc, W. T.; C. L. Moore, P. A. Zaphyr. Application of a Digital Computer to the Design of Power Transformers to Specification. (56-79).......134–7; disc. 138

Dunfee, Bernadine L. An A-C Kelvin Bridge for the Audio Frequency Range. (56-25)..123–7; disc. 128

E

Easterlin, Philip R. Control System for Integrated Data Processing (56-655).......421–7

Eberhart, E. K.; F. J. Hallenbeck, E. H. Perkins. Circuit and Equipment Descriptions of Type-P1 Carrier System. (56-242).......195–204

Ellert, F. J. Dynamic Behavior of a 3-Phase Magnetic Amplifier. (56-714).......554–60; disc. 560

Ellis, H. M.; J. W. Phelps, C. L. Roach, R. E. Treen. Co-ordinated Protection for Open-Wire Joint-Use—Ontario Tests. (56-238).......223–34; disc. 235

Emling, J. W. Disc.......732

Emling, J. W.; E. T. Mottram, R. J. Halsey, R. G. Griffith. Transatlantic Telephone Cable System—Planning and Over-all Performance.......842–50

Eppes, J. B., Jr.; K. S. Johnson. Transmission Rating of Telephone Sets. (56-781).....646–52; disc. 652

F

Farmanfarma, G. Analysis of Linear Sampled-Data Systems with Finite Pulse Width: Open Loop. (56-982).......808–19

Farrar, H. K.; J. L. Maxwell. Automatic Dispatch System for Half-Duplex Teletypewriter Lines. (56-646).......441–5

Fenaroli, A. L.; J. F. Ringelman. The Simple Reactor Circuit; Its Operation and Mode Transition. (56-750).......660–8; disc. 668

Feth, George G. Disc.......305, 636

Fischell, R. E. Magnetic Effects of Compressional Stress at Low Field Intensities. (56-133).148–51; disc. 151

Fischer, H. B.; A. W. Lebert, M. C. Biskeborn. Cable Design and Manufacture for the Transatlantic Submarine Cable System.......907–17

Fisher, W. C. Microwave for Telephone Companies.......538–42

Flanigan, J. M. Disc.......214

Fraresso, M. Disc.......235

Fraser, J. M.; H. A. Lewis, R. S. Tucker, G. H. Lovell. System Design for the North Atlantic Link...850–64

Friedlaender, F. J. Flux Reversal in Magnetic Amplifier Cores. (56-219).......268–76; disc. 278

G

Gast, Raymond W. Field Experience with the A2A Video System. (56-972).......710–16

Gaudet, G.; J. W. Noyes, S. Bonneville. Development of Transcontinental Communications in Canada. (56-244).......342–52

Geyger, W. A. Measurements on High-Speed Magnetic Servo Amplifiers with 2-Phase Motor Load. (56-218).......242–8

Gleichmann, T. F.; A. H. Lince, M. C. Wooley, F. J. Braga. Repeater Design for the North Atlantic Link.......865–77

Gohar, M. Kamal. A New Method for Treating Nonlinear Problems with Applications to Iron-Cored Oscillatory Circuits:
Part I. Subharmonics Do Not Exist. (56-983)..771–6

Gramels, J. Disc.......251

Gray, T. S.; R. H. Spencer. Transistor Voltage Regulator. (56-41).......15–17

Gremer, Charles E. An Analysis and Design of the Ferroelectric Resonant Trigger Pair. (56-119)....404–07

Griemsmann, John W. E. Disc.......674

Griffith, R. G.; E. T. Mottram, R. J. Halsey, J. W. Emling. Transatlantic Telephone Cable System—Planning and Over-all Performance.......842–50

Grim, W. M., Jr. Disc.......752

Gutzwiller, F. W. Rating and Application of Germanium and Silicon Rectifiers. (56-934)......753–

H

Halacsy, Andrew A. Disc.......593

Hallenbeck, F. J.; E. K. Eberhart, E. H. Perkins. Circuit and Equipment Descriptions of Type-P1 Carrier System. (56-242).......195–204

Halsey, R. J.; J. B. Bampton. System Design for the Newfoundland–Nova Scotia Link.......917–27

Halsey, R. J.; E. T. Mottram, J. W. Emling, R. G. Griffith, Transatlantic Telephone Cable System—Planning and Over-all Performance.......842–50

Hamer, Howard. Optimum Linear-Segment Function Generation. (56-696).......518–20

Harp, M. C.; M. H. Kebby. A 72-Channel Radio System for Toll Telephone Service. (56-16).......113–19; disc. 119

Harris, F. K. Galvanometer Efficiency as a Design Parameter. (56-60).......176–80; disc. 180

Harris, W. P.; I. L. Cooter. The Use of an A-C Bridge to Measure Core Loss at High Inductions. (56-26).......252–5; disc. 256

Hayward, H. N. Disc.......127, 180

Heffner, W. W.; H. A. Lamb. Repeater Production for the North Atlantic Link.......877–89

Henkels, Herbert W. The Fused Silicon Rectifier. (56-975).......733–46

Hersey, R. E.; J. W. Dehn. Recent New Features for the No. 5 Crossbar Switching System (56-667).......457–66

Higgins, Thomas J.; D. K. Reitan. Accurate Determination of the Capacitance of a Thin Rectangular Plate. (56-980).......761–6; disc. 766

Hoagland, Albert S. Magnetic Data Recording Theory: Head Design. (56-695).......506–12

Hoft, R. G. Inverse-Time Underfrequency Relay. (56-738).......625–9

Holden, W. H. T. Disc.......710

Holmes, M. F.; J. O. McNally, G. H. Metson, E. A. Veazie. Electron Tubes for the Transatlantic Cable System.......898–907

Honnell, P. M.; R. E. Horn. Analogue Computer Synthesis and Error Matrices. (56-49).......26–32

Hooper, Edward T.; A. D. Krall. The Operation of the Self-Balancing Magnetic Amplifier. (56-204).79–84

Horgan, James D. Disc.......766

Horn, R. E.; P. M. Honnell. Analogue Computer Synthesis and Error Matrices. (56-49).......26–32

Horowitz, I. Disc.......171

Hosford, J. A. The Development of Automatic Manufacturing Facilities for Reed Switches. (56-663).......496–500

Hovgaard, O. M. Capability of Sealed Contact Relays. (56-658)......466–8

Howard, J. D., Jr. Application of the Type-P1 Carrier System to Rural Telephone Lines. (56-241)......205–13; disc. 214

Howe, G. H. Dynamax, a New Crystal and Domain-Oriented Magnetic Core Material. (56-681)......548–50; disc. 551

Howell, S. N.; W. Derganc. Two New Total Radiation Pyrometers. (56-98)......360–7

Huang, Chaang; M. Marshall, B. H. White. Field Effect Transistor Applications. (56-252)......323–9

Huang, Chaang; E. Slobodzinski. Power Transistor Switching Circuit. (56-251)......290–6

I

Ingersoll, E. R.; M. F. Thompson, R. R. Trautwein. Magnetic-Pulse-Generator Practical Design Limitations. (56-728)......786–92

J

Jack, J. S.; W. H. Leech, H. A. Lewis. Route Selection and Cable Laying for the Transatlantic Cable System......944–55

Jensen, H. M.; V. H. Kraybill. An Instrument to Record Voltage Flicker. (56-797)......527–30

Joel, A. E. Electronics in Telephone Switching Systems. (56-668)......701–09; disc. 710

Johnson, J. R.; H. H. Anderson. Numerical Integration of Differential Equations on the 704 EDPM. (56-697)......569–74

Johnson, K. S.; J. B. Eppes, Jr. Transmission Rating of Telephone Sets. (56-781)......646–52; disc. 652

K

Kajihara, Hitoshi H. Design of Wide-Band R-F Transformers Utilizing a Synthesized Equivalent Network. (56-970)......802–05

Kapp, R. O. Disc......769

Katz, D. A Magnetic Amplifier Switching Matrix. (56-205)......236–41

Kebby, M. H.; M. C. Harp. A 72-Channel Radio System for Toll Telephone Service. (56-16)......113–19; disc. 119

Kelly, Mervin J.; Sir G. Radley. Transatlantic Communications: A Historical Résumé......841–2

Kelly, M. J.; J. E. Wallace. The Analytical Design and Evaluation of Electromagnets. (56-787)......675–80

Kelly, R.; J. F. P. Thomas. Power-Feed System for the Newfoundland–Nova Scotia Link......938–43

Kennedy, Ralph. Color-Television-System Performance Requirements. (56-271)......653–9

Kessel, A. A.; R. S. Norman, R. A. Pfuntner. Non-tracking Organic Insulations. (56-699)......257–61

Klemm, G. H. Automatic Protection Switching for TD-2 Radio System. (56-698)......520–7

Klich, R. H.; J. B. Booth. Simplified Printing Telegraph Switching and Integrated Data Processing (56-28)......11–14

Knapp, H. M. Design Features of Bell System Wire Spring Relays. (56-662)......482–6

Kochen, M.; S. Y. Wong. Automatic Network Analysis with a Digital Computation System. (56-156)......172–5; disc. 176

Krall, Albert D.; E. T. Hooper. The Operation of the Self-Balancing Magnetic Amplifier. (56-204).79–84

Kraybill, V. H.; H. M. Jensen. An Instrument to Record Voltage Flicker. (56-797)......527–30

Krengel, H. K.; R. S. Caruthers, W. S. Chaskin. A New 4-Channel Open-Wire Carrier System Employing Transistors. (56-245)......314–22

Kuck, R. G. Microwave Facilities with Built-In Reliability. (56-659)......438–41

Kuhn, R. B.; D. Levine. Method of Measuring Spot Size of Cathode-Ray Tube. (56-265).357–9; disc. 359

Kusters, N. L.; W. J. M. Moore, Disc......593

L

Lamb, H. A.; W. W. Heffner. Repeater Production for the North Atlantic Link......877–89

Lanigan, S. J.; T. J. McDonough. Traffic Considerations in Community Dial Offices. (56-645)......543–7; disc. 548

Lebert, A. W.; H. B. Fischer, M. C. Biskeborn. Cable Design and Manufacture for the Transatlantic Submarine Cable System......907–17

Lee, C. Y.; W. H. Chen. Several Valued Combinational Switching Circuits. (56-162)......278–83

Lee, T. H.; L. P. Winsor. Properties of a D-C Arc in a Magnetic Field. (56-132)......143–8; disc. 148

Leech, W. H.; J. S. Jack, H. A. Lewis. Route Selection and Cable Laying for the Transatlantic Cable System......944–55

Levi, Leo. Synchronization Accuracy Obtainable with Multiplier Phototubes. (56-759)......603–06

Levine, D.; R. B. Kuhn. Method of Measuring Spot Size of Cathode-Ray Tube. (56-265)......357–9; disc. 359

Lewis, H. A.; J. S. Jack, W. H. Leech. Route Selection and Cable Laying for the Transatlantic Cable System......944–55

Lewis, H. A.; R. S. Tucker, G. H. Lovell, J. M. Fraser. System Design for the North Atlantic Link...850–64

Lince, A. H.; T. F. Gleichmann, M. C. Wooley, F. J. Braga. Repeater Design for the North Atlantic Link......865–77

Littmann, Martin F. Disc......551

Lovell, G. H.; H. A. Lewis, R. S. Tucker, J. M. Fraser. System Design for the North Atlantic Link...850–64

Lowrance, J. L.; J. D. Dolan. Diode-Shunting in Magnetic Amplifiers. (56-749)......619–24

Lynn, G. E. Disc......554

Lynn, G. E.; J. F. Ringelman. A Signal-Discriminating Magnetic Amplifier. (56-226)......97–102

Lyons, Lambert F. Magnetic Reduction Transistor Power Supply. (56-756)......643–5

M

MacAdam, W. H. Disc......652

Malti, M. G.; H. H. Sun. Synthesis of Transfer Functions with Poles Restricted to the Negative Real Axis into Two Parallel R-C Ladders and an Ideal Transformer. (56-126)......165–71

Marihart, D. J.; D. L. Wylie. Southwest Oregon Radio System for the Bonneville Power Administration. (56-700)......531–8

Marks, G. W.; D. L. Waidelich, L. A. Monson. Investigations Concerning Polarization in Barium Titanate Ceramics. (56-682)......469–77

Marshall, Melvin; C. Huang, B. H. White. Field Effect Transistor Applications. (56-252)......323–9

Mathwich, H. R.; E. D. Nuttall, J. E. Pitman, A. M. Randolph. Propagation Test on 955.5 Mc, 1,965 Mc, and 6,700 Mc. (56-968)......685–91

Maxwell, J. L.; H. K. Farrar. Automatic Dispatch System for Half-Duplex Teletypewriter Lines. (56-646)......441–5

McDonough, T. J.; S. J. Lanigan. Traffic Considerations in Community Dial Offices. (56-645)......543-7; disc. 548

McDuffie, G. E., Jr.; C. F. Pulvari. Signals from Switched Ferroelectric Memory Capacitors. (56-755)......681–5

McKean, A. L.; E. J. Merrell, J. Arbuthnott, Jr. Styroflex Aluminum-Sheathed Air-Dielectric Cable. (56-803)......669–74; disc. 674

McMullin, N. D. Disc......420

McMurray, William. Magnetic Frequency Multipliers and Their Rating:
  Part I. Frequency Triplers. (56-229)......384–90

McNally, J. O.; G. H. Metson, E. A. Veazie, M. F. Holmes. Electron Tubes for the Transatlantic Cable System......898–907

McStay, Frank W. Disc......223

Merrell, E. J.; A. L. McKean, J. Arbuthnott, Jr. Styroflex Aluminum-Sheathed Air-Dielectric Cable. (56-803)......669–74; disc. 674

Meszar, J. The Full Stature of the Crossbar Tandem Switching System. (56-665)......486–96

Meszaros, G. W.; H. H. Spencer. Power-Feed Equipment for the North Atlantic Link......890–8

Metson, G. H.; J. O. McNally, E. A. Veazie, M. F. Holmes. Electron Tubes for the Transatlantic Cable System......898–907

Middendorf, W. H. Graphical Determination of the Fourier Series Coefficients. (56-684)......478–82

Miller, J. B.; J. L. Asmuth, C. H. Davidson, D. S. Noble, A. K. Scidmore. The Wisconsin Integrally Synchronized Computer—A University Research Project. (55-742)......330–8

Milnes, A. G. Transistor Power Amplifiers with Switched Mode of Operation. (56-94)......368–72

Molnar, Imre. Recent Developments in 4-Wire Switching of Long Distance Telephone Circuits. (56-631)......722–32; disc. 732

Molnar, Imre. Disc......547

Monson, L. A.; G. W. Marks, D. L. Waidelich. Investigations Concerning Polarization in Barium Titanate Ceramics. (56-682)......469–77

Moore, C. L.; W. T. Duboc, P. A. Zaphyr. Application of a Digital Computer to the Design of Power Transformers to Specification. (56-79)...134–7; disc. 138

Moore, W. J. M.; N. L. Kusters. Disc......593

Morris, J. A.; H. C. Thomas. Measuring 1,800,000-Kw Demand. (56-104)......95–7

Mortimer, Harry T. Disc......400

Moses, J. M.; D. I. Cone. Disc......214

Mottram, E. T.; R. J. Halsey, J. W. Emling, R. G. Griffith. Transatlantic Telephone Cable System—Planning and Over-all Performance......842–50

N

Neal, J. P. A Phasor Method of Nonlinear Network Analysis. (56-804)......630–6

Noble, D. S.; J. L. Asmuth, C. H. Davidson, J. B. Miller, A. K. Scidmore. The Wisconsin Integrally Synchronized Computer—A University Research Project. (55-742)......330–8

Noonan, Miss N. M.; S. B. Williams. A New Five-Digit Plugboard System for Card-Programmed Calculator. (56-143)......67–71

Norman, R. S.; R. A. Pfuntner, A. A. Kessel. Non-tracking Organic Insulations. (56-699)......257–61

Noyes, J. W.; G. Gaudet, S. Bonneville. Development of Transcontinental Communications in Canada. (56-244)......342–52

Nuttall, E. D.; H. R. Mathwich, J. E. Pitman, A. M. Randolph. Propagation Test on 955.5 Mc, 1,965 Mc, and 6,700 Mc. (56-968)......685–91

O

O'Brien, Joseph A. Cyclic Decimal Codes for Analogue to Digital Converters. (56-21)......120–2

P

Palmquist, T. F. Multiunit Neutralizing Transformers. (56-929)......717–21

Patton, H. W. The Pulse Stretch Coupling Circuit. (56-30)......377–9

Perkins, E. H.; E. K. Eberhart, F. J. Hallenbeck. Circuit and Equipment Descriptions of Type-P1 Carrier System. (56-242)......195–204

Perry, Donald B.; H. Bavley, B. S. Swezey. Static Elimination on Bell System Teletypewriters. (56-106)......92–4

Pfuntner, R. A.; R. S. Norman, A. A. Kessel. Non-tracking Organic Insulations. (56-699)......257–61

Phelps, J. W.; H. M. Ellis, C. L. Roach, R. E. Treen. Co-ordinated Protection for Open-Wire Joint Use—Ontario Tests. (56-238)......223–34; disc. 235

Piper, W. W. A Digital Computer Program for Determining Atomic-Wave Functions. (56-155).152–5

Pipes, Louis A. Computation of the Impedances of Nonuniform Lines by a Direct Method. (56-685)......551–4

Pitman, J. E.; H. R. Mathwich, E. D. Nuttall, A. M. Randolph. Propagation Test on 955.5 Mc, 1,965 Mc, and 6,700 Mc. (56-968)......685–91

Polk, Charles. Standing-Wave Ratio of Inaccessible Load. (56-32)......9–11

Pulvari, C. F.; G. E. McDuffie, Jr. Signals from Switched Ferroelectric Memory Capacitors. (56-755)......681–5

Pyne, Insley, B. Linear Programming on an Electronic Analogue Computer. (56-147)......139–42; disc. 143

R

Rabins, L. Transformer Reactance Calculations with Digital Computers. (56-59)......261–7; disc. 267

Radley, Sir Gordon; M. J. Kelly. Transatlantic Communications: A Historical Résumé......841–2

Ralston, Oliver C. Practical Applications of Electrostatic Phenomena to Particulate Matter. (56-114)......155–9

Ramsauer, O. Disc......530

Randolph, A. M.; H. R. Mathwich, E. D. Nuttall, J. E. Pitman. Propagation Test on 955.5 Mc, 1,965 Mc, and 6,700 Mc. (56-968)......685-91

Rea, W. T. The Communication Engineer's Needs in Information Theory. (56-653)......805–08

Reitan, Daniel K.; T. J. Higgins. Accurate Determination of the Capacitance of a Thin Rectangular Plate. (56-980)......761–6; disc. 766

Reza, F. M. Some Mathematical Properties of Root Loci for Control System Design. (56-125)...103–08

Rice, James W. Manufacture of Wire Spring Relays for Communication Switching Systems. (56-650)......513–18

Rice, S. O. A First Look at Random Noise. (56-73)......128–31

Ringelman, J. F.; A. L. Fenaroli. The Simple Reactor Circuit; Its Operation and Mode Transition. (56-750)......660–8; disc. 668

Ringelman, J. F.; G. E. Lynn. A Signal-Discriminating Magnetic Amplifier. (56-226)......97–102

Roach, C. L.; H. M. Ellis, J. W. Phelps, R. E. Treen. Co-ordinated Protection for Open-Wire Joint Use—Ontario Tests. (56-238)......223–34; disc. 235

S

Sachs, H. M.; A. L. Albin. Design of Electronic Equipment for Radio-Interference Reduction. (56-257)......352–7

Sack, E. A.; P. N. Wolfe. Disc......288

Salihi, Jalal T. Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop Core Material in Series with Capacitors. (56-10)......296–305; disc. 305

Vol.
Pt. 1
1956

Schaefer, D. H.; R. L. Van Allen. Transcendental Function Analogue Computation with Magnetic Cores. (56-134)..................160–4; disc. 165
Schiewe, A. J.; K. Chen. A Single-Transistor Magnetic-Coupled Oscillator. (56-207)................396–9
Schiewe, A. J.; K. Chen. Disc.......................164
Schmitz, N. L.; T. Bernstein. Alternating Current Control of the Half-Wave Bridge Magnetic Amplifier. (56-626)................................433–8
Schwartz, L. S. Principles of Noise Reduction in Communication Channels. (56-74)..............44–50
Scidmore, A. K.; J. L. Asmuth, C. H. Davidson, J. B. Miller, D. S. Noble. The Wisconsin Integrally Synchronized Computer—A University Research Project. (55-742)...........................330–8
Shevel, W. Lee; E. Wingrove, L. Depian. Hysteresis Loops in Dielectric Amplifiers. (56-206)..................................283–8; disc. 289
Shilliday, T. S. High-Temperature Selenium Rectifiers—A Survey of Manufacturers' Data. (56-77)..................................248–51; disc. 251
Shimanski, E. J.; W. J. Degnan. A Field Survey of Transformer Oil Quality. (56-805).575–9; disc. 579
Slobodzinski, Edwin; C. Huang. Power Transistor Switching Circuit. (56-251)................290–6
Slothower, J. C.; E. G. Albrecht, A. E. Dietz, E. W. Christoferson. Co-ordinated Protection for Open-Wire Joint Use—Minneapolis Tests. (56-239)..................................217–22; disc. 223
Smith, C. G. Disc..................................148
Smith, D. H. Power Supplies for the Type-P1 Carrier System. (56-243)..................191–4; disc. 194
Smith, J. J. Disc..................................769
Spencer, G. R. Disc...............................359
Spencer, H. H.; G. W. Meszaros. Power-Feed Equipment for the North Atlantic Link..........890–8
Spencer, R. H.; T. S. Gray. Transistor Voltage Regulator. (56-41)............................15–17
Spitzer, Charles F. The Ferroresonant Trigger Pair: Analysis and Design. (56-584)............407–16
Sproul, P. L. A Video Visual Measuring Set with Sync Pulses. (56-673)........................427–32
Sterling, R. F. Arc-Resistant Molding Materials and Finishes. (55-755).........................776–80
Storer, James E. Impedance of Thin-Wire Loop Antennas. (56-692).........................606–18
Sun, Hun Hsuan; M. G. Malti. Synthesis of Transfer Functions with Poles Restricted to the Negative Real Axis into Two Parallel R-C Ladders and an Ideal Transformer. (56-126)..............165–71
Swezey, B. S.; H. Bavley, D. B. Perry. Static Elimination on Bell System Teletypewriters. (56-106)..92–4

T

Tate, L. A.; C. R. Doty. A Data Transmission Machine. (56-985)...............................600–03
Thomas, H. C.; J. A. Morris. Measuring 1,800,000-Kw Demand. (56-104)..........................95–7
Thomas, J. F. P.; R. Kelly. Power-Feed System for the Newfoundland–Nova Scotia Link..........938–43
Thompson, M. F.; R. R. Trautwein, E. R. Ingersoll. Magnetic-Pulse-Generator Practical Design Limitations. (56-728)..........................786–92
Tostanoski, B. M. The Analysis of Sampled Data Servomechanisms Performed on the IBM Type 650. (56-680)....................................446–50
Trautwein, R. R.; M. F. Thompson, E. R. Ingersoll. Magnetic-Pulse-Generator Practical Design Limitations. (56-728)..........................786–92
Treen, R. E.; H. M. Ellis, J. W. Phelps, C. L. Roach. Co-ordinated Protection for Open-Wire Joint Use—Ontario Tests. (56-238).........223–34; disc. 235
Trousdale, Robert B. The Symmetrical Transistor as a Bilateral Switching Element. (56-585).....400–03
Tucker, R. S.; H. A. Lewis, G. H. Lovell, J. M. Fraser. System Design for the North Atlantic Link...850–64

V

Vallese, L. M. Optimum Design of Common-Emitter Transistor Audio Amplifiers. (55-667)......391–5
Vallese, L. M. Temperature Stabilization of Transistor Amplifiers. (56-276)....................379–84
Van Allen, R. L. Disc..............................400
Van Allen, R. L.; D. H. Schaefer. Transcendental Function Analogue Computation with Magnetic Cores. (56-134)..................160–4; disc. 165
Van Haste, W. Application of Statistical Technique to Electron Tubes for Use in a 4,000 Mile Transmission System. (56-86)..........................50–4
Van Nice, R. I. Magnetic Logic Circuit Control System Design Considerations. (56-727)......595–600
Veazie, E. A.; J. O. McNally, G. H. Metson, M. F. Holmes. Electron Tubes for the Transatlantic Cable System..........................898–907

W

Wahl, R. E. Direct Water-Cooled Germanium Power Rectifier. (56-976).......................832–41
Waidelich, D. L.; G. W. Marks, L. A. Monson. Investigations Concerning Polarization in Barium Titanate Ceramics (56-682)...............469–77
Wakeman, C. B.; F. J. Beck. Superposed Magnetic Fields in Materials with Rectangular Hysteresis Loops. (56-717)..........................562–9
Walker, D. C.; R. A. Brockbank, V. G. Welsby. Repeater Design for the Newfoundland–Nova Scotia Link....................................927–38
Wall, Robert E., Jr. Some of the Engineering Aspects of the Machine Translation of Language. (56-693)............................580–4; disc. 585
Wallace, J. E.; M. J. Kelly. The Analytical Design and Evaluation of Electromagnets. (56-787)....675–80
Wallenstein, G. D. Telephone Carrier Frequency Networks. (56-783)..........................691–7
Warburton, F. W. Disc.............................531
Weber, Laurance A. Influence of Noise on Telephone Signaling Circuit Performance. (56-784)...636–42
Welsby, V. G.; R. A. Brockbank, D. C. Walker. Repeater Design for the Newfoundland–Nova Scotia Link....................................927–38
Weppler, H. E.; W. R. Bullard. Co-ordinated Protection for Open-Wire Joint Use—Present Trends. (56-237)..................................215–16
Werres, C. O. Disc.................................128
Wetherill, L. Disc.................................579
Weyrick, Robert C. Transistor Analogue Computing Amplifiers for Flight Simulators. (56-261)..338–42
White, B. H.; C. Huang, M. Marshall. Field Effect Transistor Applications. (56-252)..........323–9
Williams, S. B.; P. A. Abetti. Disc.................137
Williams, S. B.; Miss N. M. Noonan. A New Five-Digit Plugboard System for Card-Programmed Calculator. (56-143)..........................67–71
Williamson, J. W. Electrical Formulas in Inch Units of Length. (56-131)..........................61–4
Wingrove, Earl; L. Depian, W. L. Shevel. Hysteresis Loops in Dielectric Amplifiers. (56-206)..................................283–8; disc. 289
Winsor, L. P.; T. H. Lee. Properties of a D-C Arc in a Magnetic Field. (56-132)........143–8; disc. 148
Wolfe, Philip. Disc................................142
Wolfe, P. N.; E. A. Sack. Disc.....................228
Wong, S. Y.; M. Kochen. Automatic Network Analysis with a Digital Computation Symem. (56-156)..................................172–5; disc. 176
Wood, E. H. Physical Response Requirements of Pressure Transducers for the Reproduction of Physiological Phenomena. (56-81)...............32–40
Woodson, H. H. Full-Wave Bridge Magnetic Amplifiers with Inductive Loads. (56-9)............7–9
Wooley, M. C.; T. F. Gleichmann, A. H. Lince, F. J. Braga. Repeater Design for the North Atlantic Link....................................865–77
Wygant, P. W. Conversion of Studio Lighting from Black and White to Color Television. (56-76)..75–8
Wylie, D. L.; D. J. Marihart. Southwest Oregon Radio System for the Bonneville Power Administration. (56-700)....................................531–8

Y

Yngve, Victor H. The Technical Feasibility of Translating Languages by Machine. (56-928).....792–7
Young, Leo. Electrical Units and Dimensions. (56-984)..........................767–8; disc. 770
Yu, Yao-Nan. The Impedance Tensor of the General Machine. (56-122)......................181–7

Z

Zaborszky, John. Integrated s-Plane Synthesis Using 2-Way Root Locus. (56-981)..............797–801
Zaphyr, P. A.; C. L. Moore, W. T. Duboc. Application of a Digital Computer to the Design of Power Transformers to Specification. (56-79)..134–7; disc. 138

D-C Graphical Analysis of Junction Transistor Flip-Flops

T. R. BASHKOW
NONMEMBER AIEE

THE D-C analysis and design of nonlinear circuits may be greatly facilitated by graphical means. A graphical approach, if it is simple enough, seems to have the happy faculty of providing understanding of existing circuit operation and hints for better or different designs. This has been amply illustrated recently by the wide use of a negative resistance curve for point contact trigger circuits.

A graphical analysis of the circuit of Fig. 1 has therefore been developed. This circuit, or variations thereof, has been analyzed previously[1] by straightforward algebraic methods assuming one transistor is ON and the other OFF. The resulting set of relations among the various parameters is then used to insure that the initial assumptions are satisfied.

The analysis to be presented makes no initial assumptions as to the d-c state of the circuit. Instead, a modification of the conventional graphical technique of plotting a load line on a set of device characteristics is used. In this case, however, two load lines are required, except that only one is needed for a completely symmetrical circuit. Since the load lines include all possible d-c operating points they must include the location of the d-c stable points as well. A variation in any of the circuit parameters produces observable changes in the stable points and thus bistability or, if desired, monostability can be attained.

The method is not exact, because a linearization of transistor characteristics is made, but experimental checks on the results have been quite good. As with all graphical methods, this one shows a change in resistance or voltage as a change of slope or intercept of straight lines and is therefore amenable to use as a design technique; see example in the Appendix.

General Description of Method

The following analysis can be generally applied to any circuit of this type whether the active element be transistor or vacuum tube, but for illustration only *p-n-p* junction transistors will be used. The approach is to consider one part of the circuit as isolated and the rest of the circuit as a single load. (The labeling of Fig. 1 reflects this in that R_2, R_c, R_1 and the base-to-ground resistance of T_b are the load for T_a. The load for T_b is R_2', R_c', R_1', and the base-to-ground resistance of T_a.) The isolated circuit is that shown in Fig. 2(A). Plotting V_2 versus I_2 the usual collector characteristics of T_a, is obtained, modified by R_2, E and V_{cc}, shown in Fig. 2(B). The conventional drive current, I_b', is here labeled in terms of I_1'. If $R_1' \rightarrow \infty$ these, of course, are identical.

The load for the circuit of Fig. 2 is given in Fig. 3(A). Note that $I_2=-I_1$. This load consists of R_c in series with the parallel combination of R_1 and the input resistance of T_b. When T_b is ON, the load is $\cong R_c$; when T_b is OFF the load is $\cong R_c+R_1$. In Fig. 3(B), the dashed line represents the parallel combination of R_1 and the base-to-ground resistance of T_b. The solid line, the actual load line, results when R_c is added graphically.

If the load line of Fig. 3(B) is plotted on Fig. 2(B), Fig. 4(A) results and, assuming the same procedure for T_b the result is Fig. 4(B).

The load lines of Fig. 4 consist of all possible d-c operating points for this circuit. The end points of the load lines, points 1 and 2, are the two stable points required for a bistable circuit. To indicate that this is so, consider that T_b is OFF, operating at point 1B. Reading the current co-ordinate of this point shows that $-I_1'$ is sufficient to turn T_a ON. Therefore T_a must be ON at the point *1a*. Reading the current co-ordinate of this point it is seen that $-I_1$ is not sufficient to turn T_b ON. Therefore, T_b must be OFF at the point *1b*. This is the starting point and the circuit is stable with T_b OFF, T_a ON.

Similarly it is seen that a second stable point must be at 2 with T_b ON, T_a OFF. Later, it will be shown that one of the stable points may lie in the active, rather than the saturated, region. A third possible operating point requires both transistors to be in their active region. (All three operating points are d-c stable. It is fairly obvious that the two d-c stable points of Fig. 4 are also dynamically stable since the OFF transistor opens the feedback loop. Dynamic stability of the third point must be investigated by methods outside the scope of this presentation.)

In the next sections specific information is given concerning the modified collector family and the d-c load line. It will be shown that these can easily be drawn with a knowledge of circuit parameters and a few transistor parameters.

Modified Collector Family

For the *p-n-p* transistor circuit of Fig. (5A)

$$I_2=\frac{V_2-V_{cc}}{R_2}+I_c \qquad (1)$$

may be written as the equation for the currents at the collector node.

For $I_c=0$, T_a is cut off, and $I_2=(V_2-V_{cc})/R_2$ which is plotted as the line of slope R_2 connecting the points $I_2=-V_{cc}/R_2$ and $V_2=V_{cc}$ of Fig. 5(B). Since I_c does not go to zero under normal conditions this cutoff line represents a limit rather than an actuality, but is approached very closely with a slightly positive value of I_b'. (As would be expected, the minimum value of I_c depends very strongly on I_{co} and as $|I_{co}|$ increases, the actual cutoff line moves further from this limit.)

For $V_2=E$, the transistor is in saturation and presents a low resistance r. For this condition, if, in addition, $I_c=0$, the intersection is obtained of saturation and cutoff lines at $V_2=E$, $I_2=(E-V_{cc})/R_2$, as shown in Fig. 5(B).

The lines of constant I_1' in the active region of Fig. 5(B) may also be considered to be of slope R_2. This is, of course, not strictly true because of the finite collector resistance, but experimental results indicate that it is a good approximation. The reason is twofold: 1. R_2 shunts the collector resistance $(1-\alpha)r_c$. If R_2 is much smaller than $(1-\alpha)r_c$, it is the only effective resistance; 2. If R_2 is increased, holding V_{cc} constant, the collector current I_c decreases. As shown by Early,[2] this causes an increase in r_c and allows the condition $R_2<<(1-\alpha)r_c$ to be maintained.

Paper 56-24, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted July 26, 1955; made available for printing November 10, 1955.

T. R. BASHKOW is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

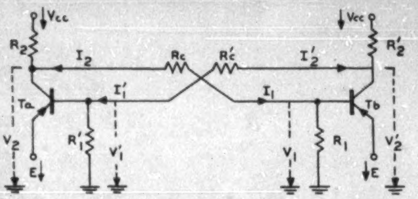

Vcc
R2
I2
Rc
R'c
I'2
Vcc
R'2
Ta
I'1
I1
Tb
V2
R'1
V'1
V1
R1
V'2
E
E

Fig. 1. Flip-flop circuit with p-n-p transistors. Only elements important at d-c are shown. Voltage arrows point in direction of positive voltage drop

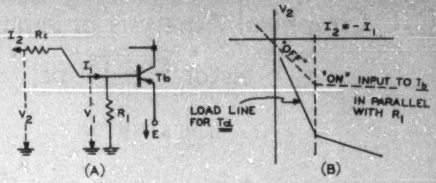

I2
Rc
I1
Tb
V2
V1
R1
E
(A)
V2
I2 = -I1
"OFF"
"ON" INPUT TO Tb
IN PARALLEL WITH R1
LOAD LINE FOR Ta
(B)

Fig. 3. Load line for T_a and portion of circuit from which it is obtained

Since the characteristic curves in the active region are straight lines of slope R_2 it is only necessary to find their intercepts, for each value of I_1'.

For the intercept at $I_2=0$ it is clear from equation 1 that $V_2=V_{cc}-I_cR_2$. Since

$$I_c=\frac{\alpha_n}{(1-\alpha_n)}I_b'+\frac{I_{c0}}{1-\alpha_n}$$

for each value of I_b', for a given transistor, the intercept can be found and, with the slope known, the corresponding line in the active region can be plotted. However, it is required that these lines be labeled in terms of I_1' and thus a relationship between I_1' and I_b' is needed. From Fig. 5 it is seen that

$$I_1'-\frac{V_1'}{R_1'}=I_b'$$

but I_b' is also dependent on V_1' so this equation must be solved graphically. It is apparent, however, that I_1' must be more negative than E/R_1' in order to overcome the emitter bias E and get the transistor out of cutoff. Increasingly negative values of I_1' move the transistor into the active region and eventually into saturation.

The modified collector family can thus be drawn from a knowledge of V_{cc}, R_2, E, and the saturation slope, which is a measurable constant for a given transistor. To label the curves in terms of I_1', it is necessary to measure I_{c0} and α_n. The relationship between I_b' and I_1' as well as an illustrative example are given in the Appendix.

Load Line

For the circuit of Fig. 6 $V_2=V_1-I_2R_c$ may be written or, since $I_2=-I_1$

$$V_2=V_1+I_1R_c \tag{2}$$

Equation 2 may be solved by plotting V_1 versus I_1 and adding I_1R_c graphically.

A linearized plot of V_1 versus I_1 for a *p-n-p* transistor is shown in Fig. 7(A) and 7(B) for two values of E. When the transistor is OFF, the base-to-ground resistance is very high and the external base resistance R_1 relates V_1 to I_1. When the transistor is ON, the low resistance r_T shunts R_1. The effect of emitter bias is simply to cause a translation of the zero bias curve of Fig. 7(A) by E along the voltage axis and E/R_1 along the current axis. The current I_T and the slope r_T are constants for a given transistor and must be measured. (That I_T is a transistor constant, independent of R_1, may seem strange until it is realized that at $I_1=I_T$, $V_1=0$. With an effective dead short from base to ground it is immaterial what resistance is connected from base to ground.) Fig. 7 shows no dependence on I_2', the current into the collector node of T_b. This is to be expected of any unilateral device. It is true of a *p-n-p* transistor as long as $I_2'<-V_{cc}/R_2'$, which keeps the collector junction in its normal state of reverse bias. This condition on I_2' holds under normal circuit operation.

Having V_1 versus I_1 the desired load line may now be obtained by adding I_1R_c to each value of V_1. Since I_1R_c is just a line of slope R_c, it may be added

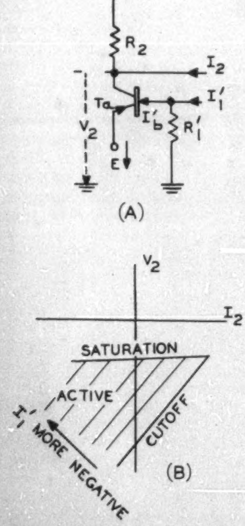

Vcc
R2
I2
I'1
Ta
I'b
R'1
V2
E
(A)
V2
I2
SATURATION
ACTIVE
CUTOFF
I'1 MORE NEGATIVE
(B)

Fig. 2. Modified collector family for T_a and portion of circuit from which it is obtained

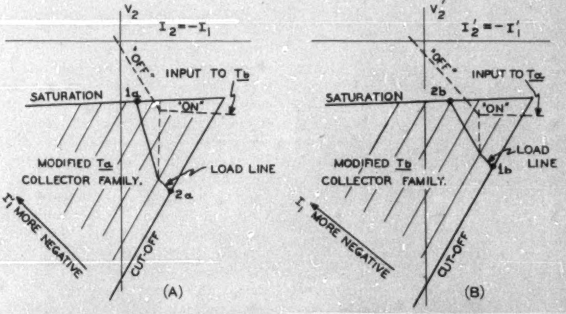

V2
I2=-I1
"OFF"
INPUT TO Tb
SATURATION
1a
"ON"
MODIFIED Ta COLLECTOR FAMILY.
LOAD LINE
2a
I1 MORE NEGATIVE
CUT-OFF
(A)
V'2
I'2=-I'1
"OFF"
INPUT TO Ta
SATURATION
2b
"ON"
LOAD LINE
MODIFIED Tb COLLECTOR FAMILY.
1b
I1 MORE NEGATIVE
CUT-OFF
(B)

Fig. 4. Load line plotted on modified collector family for each transistor

Vol. 7
Pt. 1
1956

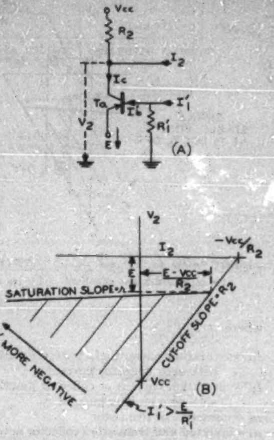


Fig. 5. Slopes and intercepts of modified collector family for T_a in terms of circuit parameters

graphically to the V_1, I_1 curve by the constructions shown in Fig. 8, which has been greatly exaggerated for clarity. The resulting broken line, shown solid in Fig. 8, is the desired load line. The construction is quite simple.

1. Pick any point on the V_1-I_1 curve.
2. At this value of V_1, on the voltage axis, draw a line of slope R_c.
3. The intersection of this line with the initial value of I_1 which was chosen gives one point on the load line.

Actually, only three points need be chosen, one point in the OFF region, one in the ON region, and the break point between them. As a check, the load line should intersect the V_1-I_1 curve at $I_1=0$, since for this case $V_2=V_1$ according to equation 2.

The modified collector curves are plotted as V_2 versus I_2 and the load line of Fig. 8 is plotted as V_2 versus $-I_2$. The latter must therefore be rotated about the voltage axis to obtain Figs. 3(B) and 4(B).

The load line can thus be drawn from a knowledge of R_1, R_c, E, and the simple construction of Fig. 8. In addition, the transistor constant I_T (which is of the order of I_{c0}) is needed and the ON base-to-ground resistance r_T, both of which must be measured; see the Appendix.

Use of Method

It should be emphasized that the simplification used in deriving the load line, e.g., that the transistor goes abruptly from ON to OFF at a critical value of I_1 is, of course, a fiction. It is obvious from the modified collector curves that there is a range of I_1 over which the transistor goes from ON to OFF. The α of the transistor is of particular importance in determining how closely the reality of transistor operation corresponds to the fiction. For precise design the method of the Appendix for relating I_1 to I_b (or I_1' to I_b') and labeling the modified collector characteristic must be followed. However, for a qualitative understanding of circuit action the simplification is quite useful and will be adhered to in the following discussion. This analysis provides results in the form of two graphs as shown in Fig. 9.

The two load lines contain all possible d-c operating points of the circuit and in particular the two end points are d-c stable points as explained earlier. However, suppose that R_1' is decreased so as obtain Fig. 10. In this case the circuit can never have a stable point with T_b OFF, T_a ON. To indicate this, assume that T_b is OFF; see Fig. 10(B). Reading the current co-ordinate of this point it is seen that $-I_1'$ is not sufficient to turn T_a ON. Therefore T_a must be OFF; see Fig. 10(A). Reading the current co-ordinate of this point it is seen that $-I_1$ is sufficient to turn T_b ON. Therefore T_b must be ON; see Fig. 10(B). However, this contradicts the initial assumption. Therefore the d-c condition of T_b OFF and T_a ON cannot exist. A similar analysis for T_b ON, and T_a OFF gives consistent results for both graphs simultaneously and is therefore the only d-c stable point. Fig. 11 shows another monostable circuit in which both transistors are OFF.

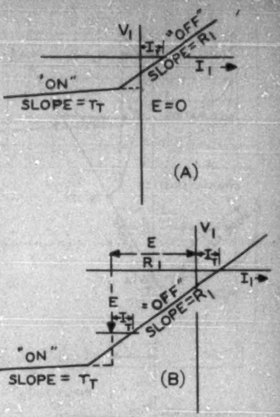


Fig. 7. Plot of voltage versus current (V_1 versus I_1), at base of T_b, for two values of emitter bias

For the following discussion a completely symmetrical circuit will be assumed, $R_1=R_1'$, $R_2=R_2'$, $R_c=R_c'$, T_a identical to T_b. In this case both graphs are identical. Only one need be drawn as in Fig. 12 which also shows the effect of a change in R_c. Assume that T_a is OFF, T_b ON. As R_c is increased, only the load line is affected. The OFF stable point then moves so that the current $-I_1$, keeping T_b ON, is decreased. In other words, T_b is coming further out of saturation. Eventually, T_b does not have enough base current to remain ON and both transistors are OFF. The two load lines whose OFF points are on either side of the dashed vertical line in Fig. 12 bracket a critical value of R_c, which allows bistability. A value of R_c slightly lower than this allows one transistor to be off

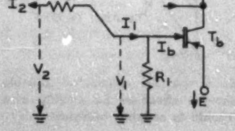


Fig. 6 (left). Load circuit for T_a

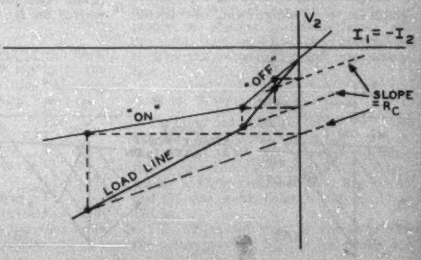


Fig. 8 (right). Graphical construction which gives load line

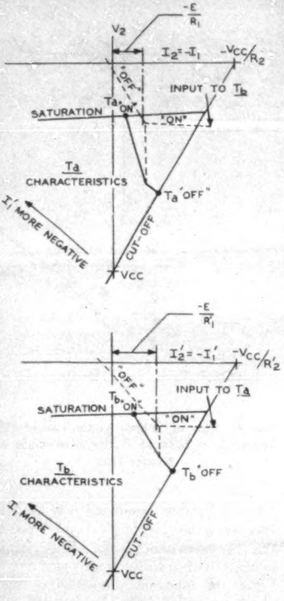

V2
-E/R1
I2=-I1
-Vcc/R2
"OFF"
INPUT TO Tb
Ta "ON"
SATURATION
"ON"
Ta CHARACTERISTICS
Ta "OFF"
I1 MORE NEGATIVE
CUT-OFF
Vcc
-E/R'1
I'2=-I'1
-Vcc/R'2
"OFF"
INPUT TO Ta
Tb "ON"
SATURATION
"ON"
Tb CHARACTERISTICS
Tb "OFF"
I1 MORE NEGATIVE
CUT-OFF
Vcc

Fig. 9. Typical graphs for circuit with two stable d-c states

and the other to be on, but not saturated. A slightly higher value of R_c does not support bistability. A study of Fig. 12 and previous figures makes it evident that increasing R_2, making E more negative, decreasing R_1, or decreasing V_{cc} causes a similar effect to that just described for increasing R_c. It is therefore possible to design a bistable circuit which does not saturate, without using any additional circuit elements. Such a design would require rather close tolerances on components, and so seems impractical, but is mentioned to indicate a possible use of these graphs.

Other information also apparent on the graph is: 1. Voltage swing of collector, 2. Current through ON transistor, and 3. Voltage at base of OFF transistor. These conditions are illustrated in Fig. 13.

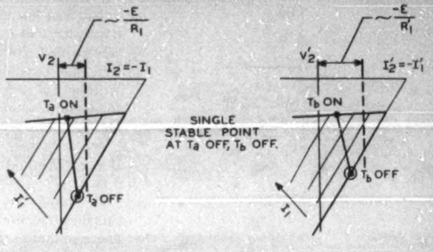

-E/R1
V2
I2=-I1
Ta ON
I1
Ta OFF
SINGLE STABLE POINT AT Ta OFF, Tb OFF.
-E/R'1
V'2
I'2=-I'1
Tb ON
I1
Tb OFF

Fig. 11 (right). Typical graphs for circuit with one stable d-c state, both transistors OFF

Some additional information, such as the effect of changes in α or I_{c0}, can also be obtained. However, these require more detailed analysis, such as that given in the example in the Appendix, and will not be given here.

Conclusions

The graphical analysis provides a picture from which the operation of a junction transistor bistable circuit can be understood by inspection. Since all circuit parameters appear explicitly, it becomes an easy matter to predict the effect of circuit variation on such things as collector voltage swing and maximum collector current. Furthermore, it is apparent whether or not the circuit values chosen will work at all, i.e., whether bistability can exist. The proper changes in parameter values to insure bistability can then be made and since the required graphs are simple to construct many possible designs can be tried in a short time.

Appendix. Relationships, Examples, and Measurements

Relationship between I₁ and I_b (or I₁′ and I_b′)

From the general expressions for alloy junction transistors given by Moll and Ebers[3] the following expression can be derived for I_b.

$$I_b=\frac{(1-\alpha_n)I_{e0}}{1-\alpha_n\,\alpha_i}\left(e^{-\frac{(V_1-E)}{0.026}}-1\right)-\frac{(1-\alpha_i)\,I_{c0}}{1-\alpha_n\,\alpha_i} \tag{3}$$

where

I_{e0} = saturation current of emitter junction with zero collector current
I_{c0} = saturation current of collector junction with zero emitter current
α_n = normal α of transistor
α_i = inverted α of transistor (collector acting as emitter, emitter acting as collector)
$kT/q \cong 0.026$ volts at room temperature

Equation 3 describes an exponential curve which can be plotted for a given transistor once the four transistor constants are known. (At the end of the Appendix some simple measurements for obtaining these transistor parameters are described.) The form is that of Fig. 14. Fig. 15 shows plots of equation 3 for two transistors with parameters as indicated and $E=0$. It is seen now that the I_T given as a constant for the V_1, I_1 approximation in the text is $-(K_1+K_2)\cong -K_2$.

The equation

$$I_1-\frac{V_1}{R_1}=I_b$$

can now be used to obtain the desired relation between I_1 and I_b. In the V_1, I_1 plane the left-hand side, I_1-V_1/R_1, is simply a line of slope $-R_1$ displaced along the current axis by I_1. The right-hand side is the previously determined curve for I_b. The graphical method of solution is therefore as shown in Fig. 16(A).

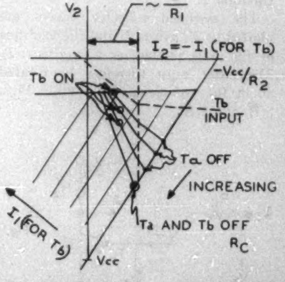

V2
-E/R1
I2=-I1 (FOR Tb)
Tb ON
-Vcc/R2
Tb INPUT
Ta OFF
INCREASING
I1 (FOR Tb)
Ta AND Tb OFF
RC
Vcc

Fig. 12. Graph for symmetrical bistable circuit showing effect on d-c stability of a change in coupling resistance

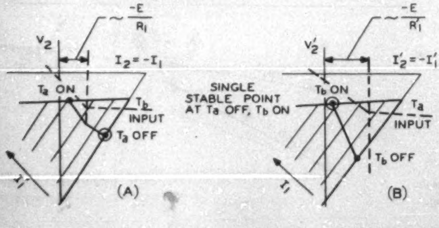

-E/R1
V2
I2=-I1
Ta ON
Tb INPUT
Ta OFF
I1
(A)
SINGLE STABLE POINT AT Ta OFF, Tb ON
-E/R'1
V'2
I'2=-I'1
Tb ON
Ta INPUT
Tb OFF
I1
(B)

Fig. 10 (left). Typical graphs for circuit with one stable d-c state, one transistor ON and the other OFF

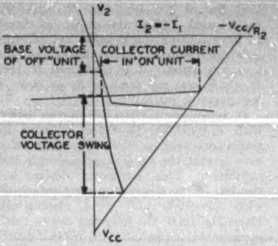


Fig. 13. Graph labeled to show transistor voltages and currents

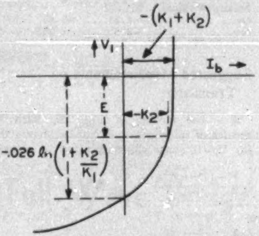


Fig. 14. General theoretical relation between base current and voltage

$$K_1=\frac{(1-\alpha_n)I_{e0}}{1-\alpha_n\alpha_i}$$

$$K_2=\frac{(1-\alpha_i)I_{e0}}{1-\alpha_n\alpha_i}$$

$$I_b=K_1\left[e^{\frac{-(V_1-E)}{0.026}}-1\right]-K_2$$

Fig. 15 (right). Theoretical base current versus voltage curves for transistors with parameters as indicated

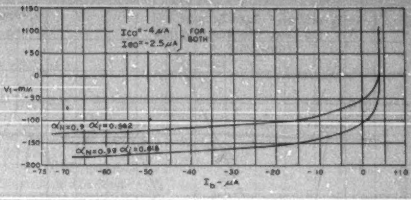
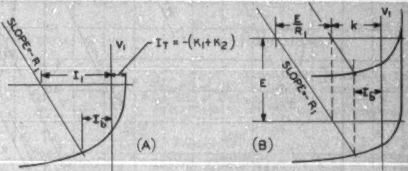


Fig. 16 (right). Graphical solution of $I_1=1/R_1=I_b$



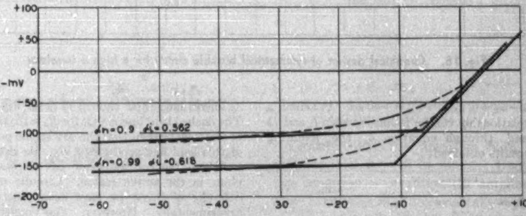


Fig. 17. Plot of V^1 versus I_1, $R_1=10{,}000\Omega$, for a high α and a low α transistor. Dotted curves are theoretical, solid curves are broken line approximations

This shows clearly why all V_1-I_1 plots go through the point I_T at $V_1=0$, regardless of the value of R_1. Fig. 16(B) shows that if $I_1=k$ for a given I_b with $E=0$, then $I_1=E/R_1+k$ for the same value of I_b. Fig. 17 shows a V_1-I_1 plot for $R_1=10{,}000\Omega$ with the curves of Fig. 15. The dashed lines are solutions of $I_1-V_1/R_1=I_b$, the solid lines are the suggested linear approximations.

Examples of Method as Design Technique

Assume a symmetrical circuit for which $V_{cc}=-5.9$ volts, $E=-1.5$ volts, $R_1=R_1'=10{,}000\Omega$ $R_2=R_2'=10{,}000\Omega$. It is desired to find a value of $R_c=R_c'$, which just insures a bistable circuit. This is not necessarily a realistic problem but will illustrate the graphical method and associated calculations.

Case I: High α Units

The load line and modified collector family are given in Fig. 18(A). The load line was constructed using $R_c=15{,}000\Omega$ and the lower solid curve of Fig. 17. This value of R_c was chosen so that the OFF stable point is just beyond the "break" point of the V_1, I_1 curve, $-I_1=-E/R_1+20\mu a=170\ \mu a$.

To check this design the lower curve is used of Fig. 15 (from which the lower solid curve of Fig. 17 was obtained). From the graphical method of the previous section the following relationship is obtained between I_1 and I_b:

I_1	I_b
+ 5.0 μa	+4.0 μa
+ 0.0 μa	+3.7 μa
− 5.0 μa	+2.7 μa
−10.0 μa	0.0 μa
−15.0 μa	−3.2 μa

To find the $I_2=0$ intercept of the lines, I_1=constant, in the active region of the modified collector family, the following relation is used

$$V_2=V_{cc}-I_cR_2=V_{cc}-\left[\frac{\alpha_n}{1-\alpha_n}I_b+\frac{I_{c0}}{1-\alpha_n}\right]R_2 \quad (4)$$

For this transistor $\alpha_n/1-\alpha_n=0.99/0.01=99$, $I_{c0}/1-\alpha_n=-400\ \mu a$. The following table summarizes the numerical results.

I_1, μa	I_c, μa	$-I_cR_2$, volts	V_2, volts
+ 5.0	− 4*	+0.04	−5.86
+ 0.0	− 34	+0.34	−5.56
− 5.0	−132	+1.32	−4.58
−10.0	−400	+4.00	−1.90
−15.0	−716	+7.16	+1.26

* It can be shown that this minimum value of $I_C=\frac{1-\alpha_I}{1-\alpha_N\alpha_I}I_{c0}$. This value determines how closely the ideal cutoff line may be approached. It is seen that an increase in $|I_{c0}|$ causes a shift of the entire collector family to the left.

These lines are shown in Fig. 18(B). It can be seen that the graphical solution for $R_c=15{,}000\Omega$ of Fig. 18(A) is indeed proper to produce bistability since the OFF stable point at $-I_1=170\ \mu a$ is sufficient to drive the ON transistor well into saturation.

Case II: Low α Units

The modified collector family and load line for $R_c=15{,}000\Omega$ are given again in Fig. 19(A). This was constructed using the upper solid curve of Fig. 17. There is only a slight difference from Fig. 18(A) and the OFF stable point is still at $-I_1=170\ \mu a$.

This design will now be checked, with the upper curve of Fig. 15 for which $\alpha_n/$

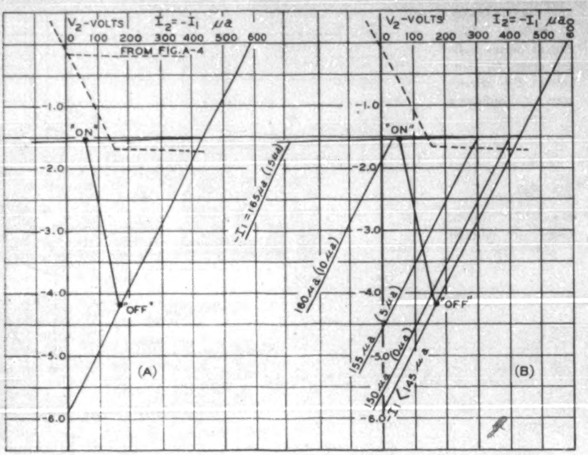


Fig. 18. Graphical design of symmetrical bistable circuit for a high α transistor

$1-\alpha_n=0$, $I_{c0}/1-\alpha_n=-40\ \mu a$. As before, a relationship can be found between I_1 and Ib and, through equation 4, the following results obtained:

I_1, μa	I_b, μa	I_c, μa	$-I_cR_2$, volts	V_1, volts
+ 5.0	+ 4.06	− 3.5	+0.04	−5.86
+ 0.0	+ 2.7	− 15.9	+0.16	−5.74
−20.0	−11.4	−142.5	+1.42	−4.48
−40.0	−29.4	−304.0	+3.04	−2.86
−60.0	−48.0	−472.0	+4.72	−1.18

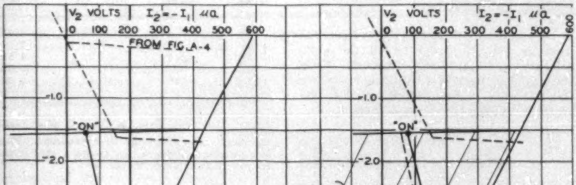


Fig. 19. Graphical design of symmetrical bistable circuit for a low α transistor

These lines are shown in Fig. 19(B). The dashed load line is that for $R_c=15{,}000\Omega$. It can now be seen that the OFF stable point of this load line produces a bistable circuit in which the ON transistor has its stable point in the active region. Circuits with behavior of this type have been observed experimentally. However, the collector voltage swing is not the maximum available and they are so sensitive as to swing very easily from one state to the other.

To obtain the maximum voltage swing and decrease trigger sensitivity, the solid load line of Fig. 19(B) for $R_c=10{,}000\Omega$, can be used. This value of R_c was chosen to have an OFF stable point at $-I_1=210\ \mu a$ which is sufficient to put the ON transistor into saturation. If it is desired to keep $R_c=15K$, one can increase R_1 or make E less negative instead.

It is apparent that this latter design, with $R_c=10{,}000\Omega$ may also be used for transistors with $\alpha_n>0.9$. The only difference is that the ON unit goes deeper into saturation. This particular design was checked in the laboratory using transistors with $\alpha_n=0.989$ and 0.944 respectively. From Fig. 19(B) the following can be read:

	V_2, volts	V_1, volts	I_1, μa
ON	−1.55	−0.8	− 75
OFF	−3.75	−1.65	−210
Measured Values			
ON	−1.55	−0.83	− 68
OFF	−3.8	−1.63	−200

Suggested Measurements on Transistor

For the circuit of Fig. 20, with the results of reference 3, it can be shown that for $V_1=0$, (base short-circuited to ground)

$$I_b=\frac{-(1-\alpha_i)I_{c0}}{1-\alpha_n\alpha_i}$$

$$I_c=\frac{I_{c0}}{1-\alpha_n\alpha_i}$$

and for $I_b=0$, (base open-circuited)

$$I_c=\frac{I_{c0}}{1-\alpha_n}$$

For these measurements the voltage at the collector should be kept at least slightly negative, −0.5 volt for a *p-n-p* transistor, otherwise the assumptions used in deriving these results are invalidated. With the relations

$$\frac{\alpha_n}{1-\alpha_n}=\frac{\dfrac{I_{c0}}{1-\alpha_n\alpha_i}-\dfrac{I_{c0}}{1-\alpha_n}}{\dfrac{-(1-\alpha_i)I_{c0}}{1-\alpha_n\alpha_i}}=\alpha'$$

or

$$\alpha_n=\frac{\alpha'}{1+\alpha'}$$

$$\alpha_i=1+\frac{\dfrac{-(1-\alpha_i)I_{c0}}{1-\alpha_n\alpha_i}}{\dfrac{I_{c0}}{1-\alpha_n\alpha_i}}$$

$$\alpha_nI_{e0}=\alpha_iI_{c0}$$

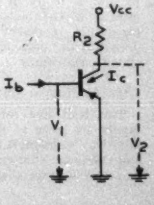


Fig. 20 (right). Circuit used for measurement of transistor parameters all transistor parameters can be obtained from the suggested measurements. (Any convenient measuring equipment can, of course, be used to obtain these parameters. This method recommended itself because two of the measurements are used directly in the calculations indicated previously. They do not depend on the accuracy of separate α and I_{c0} measurements.)

To illustrate, a set of measurements for a *p-n-p* transistor gave for $V_1=0$

$I_e=-13\ \mu a$
$I_b=+6\ \mu a$

and for $I_b=0$

$I_e=-260\ \mu a$

Therefore

$$\frac{\alpha_n}{1-\alpha_n}=\frac{-13+260}{6}=41.2$$

or

$$\alpha_n=\frac{41.2}{42.2}=0.976$$

$$\alpha_i=1+\frac{+6}{-13}=0.538$$

$$I_{c0}=-260(1-0.976)=-6.25\ \mu a$$

$$I_{e0}=\frac{0.538}{0.976}(-6.25\ \mu a)=-3.44\ \mu a$$

The circuit of Fig. 20 can also be used to obtain the slope r_T for the V_1, I_1 curves, and the saturation slope r. A current source large enough to saturate the transistor is required. Measurements on a *p-n-p* transistor gave

$V_1=-0.184$ volts for $I_b=-200\ \mu a$

$V_1=-0.167$ volts for $I_b=-100\ \mu a$

Therefore

$$r_T=\frac{0.184-0.167}{100\times 10^{-6}}=170\Omega$$

A measurement of V_2 at $I_b=-800\ \mu a$, ($I_2=0$, $E=0$) gave

$V_2=-33$ millivolts

We assume that $V_2=0$ at $I_2=-V_{cc}/R_2$. This measurement gave

$$I_2=\frac{5.9 \text{ volts}}{5,200}=1.135 \text{ milliamperes}$$

Therefore

$$r=\frac{33}{1.135}=29.1\Omega$$

References

1. JUNCTION TRANSISTOR MULTIVIBRATORS AND FLIP-FLOPS, E. W. Sard. *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 2, "Circuit Theory, 1954 National Convention," pp. 121–22.

2. EFFECTS OF SPACE CHARGE WIDENING IN JUNCTION TRANSISTORS, J. M. Early. *Proceedings, Ibid.*, vol. 40, Nov. 1952, pp. 1401–06.

3. LARGE SIGNAL BEHAVIOR OF JUNCTION TRANSISTORS, J. J. Ebers, J. L. Moll. *Proceedings, Ibid.*, vol. 42, Dec. 1954, pp. 1761–72.

Full-Wave Bridge Magnetic Amplifiers with Inductive Loads

H. H. WOODSON
ASSOCIATE MEMBER AIEE

THERE are numerous articles treating magnetic amplifiers with inductive loads.[1–4] In the majority of these papers the inductive loads are considered without so-called free-wheeling or discharge paths. When single-ended d-c output circuits are considered in the literature, the conditions are first derived for which the circuit will be unstable with no discharge paths;[1,3] then discharge paths are added to make the system stable. In the references cited, the discharge rectifiers are considered to have zero forward resistance, thereby achieving complete decoupling of the inductive load from the controlling reactors during the reset period.

The purpose of this paper is to investigate empirically the effect of discharge rectifiers, in which the forward resistance is assumed to be appreciable, on the operation of a single-ended amplifier with inductive load. An additional assumption is that the load has a time constant which is long compared to the period of the supply. The result of this study is presented in the form of a block diagram to facilitate the use of the results in feedback control system design and in design for which analogue computers are employed.

Resistive Load

The magnetic amplifier used in this study is the full-wave bridge circuit shown schematically in Fig. 1. The components for this circuit are listed in the following.

Reactors: 2-mil Orthonol $1^1/_2$ inches by $2^1/_2$ inches by 1 inch; $N_g=250$ turns no. 20 wire, $N_c=600$ turns no. 30 wire
RX = radio receptor selenium rectifiers, 26-volt cells, $1^1/_2$ inches by $1^1/_2$ inches, 6 cells in series
$R_c=2{,}050$ ohms including the resistance of $2N_c$
$R_L=190$ ohms
$L=32.4$ henrys
$\tau_L=L/(R_L+R_g)=0.153$ second
$v_g=120\sqrt{2}\sin 2\pi\, 400\, t$
R_g = apparent internal resistance of amplifier at full output = 22 ohms

Consider first the amplifier with a resistive load $L=0$ in Fig. 1. Under this condition of operation the amplifier may be represented incrementally by a gain and a first order lag[4]

$$\frac{\partial I_L}{\partial N_cI_c}=G(s)=\frac{K/R_L}{\tau_m s+1}\ \frac{\text{amperes (amp)}}{\text{amp-turn}} \qquad (1)$$

where

I_L = average load current
I_c = average control current
K = gain, volts per amp-turn
$\tau_m=(N_c^2/R_c)k_c$
R_L = load resistance
k_c = constant of the load circuit and reactors

The gain K is found at a certain output level from the steady-state transfer characteristic such as that given as curve A in Fig. 2. For this example, the output level is 0.2 amp and

$$K=0.250\ \frac{\text{volts}}{\text{amp-turn}} \qquad (2)$$

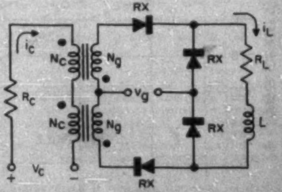


Fig. 1. Full-wave bridge magnetic amplifier circuit

Paper 56-9, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted August 2, 1955; made available for printing October 11, 1955.

H. H. WOODSON is with the Massachusetts Institute of Technology, Cambridge, Mass.

The research on which this paper is based was supported, in part, by the Department of the U. S. Air Force under Contract *AF 33(616)2669*.

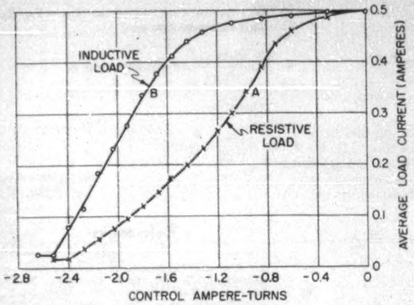


Fig. 2 (left). Steady-state transfer characteristics

The time constant τ_m is found from the frequency response curves A of Figs. 3(A) and (B)

$$\tau_m = 0.047 \text{ second} \qquad (3)$$

These constants will be used in the treatment of inductive loads to follow.

Inductive Load

An inductive load having a time constant which is very long compared to the period of the supply will have a continuous current flowing through it during the entire period of the supply voltage. Thus, during the intervals of time when both reactors are unsaturated, the load current will flow through both discharge rectifiers, as indicated by path A in Fig. 4. When one reactor is saturated, the load current will flow through only one discharge rectifier as indicated by path B or C in Fig. 4. In general, the same voltage drop v_f appears across both nonideal discharge rectifiers when they are conducting.

When the discharge rectifiers are assumed to have zero forward resistance, the voltage drops v_f disappear and the inductive load can have no effect on the unsaturated flux change in either reactor assuming that the saturated coupling between gate and control windings is negligible. Therefore, the load is decoupled from the reactors during the reset period and the incremental transfer function with inductive load becomes the product of the amplifier transfer function of equation 1 and the first-order lag of the inductive load

$$G'(s) = \frac{K/R_L}{(\tau_m s+1)(\tau_L s+1)} \ \frac{\text{amp}}{\text{amp-turn}} \qquad (4)$$

On the other hand, when the discharge rectifiers have appreciable forward resistances, the voltage drops v_f shown in Fig. 4 can affect the flux changes in the reactors. The voltage drops v_f will increase with increasing load current, and reference to Fig. 5 shows that these voltages tend to saturate the reactors. Therefore, the voltages v_f will cause a positive feedback which varies with the average load current I_L. This feedback can be represented by the feedback factor β shown in the block diagram of Fig. 5. In general, the factor β will not be a constant but will depend on the output level as well as on the circuit parameters.

The transfer function for the block diagram of Fig. 5 is

$$G''(s) = \frac{\dfrac{K'}{1-\beta K'}}{\dfrac{\tau_m \tau_L}{1-\beta K'}s^2 + \dfrac{\tau_m+\tau_L}{1-\beta K'}s + 1} \qquad (5)$$

where $K' = K/R_L$. From a plot of the steady-state transfer characteristic of the amplifier with inductive load the factor β can be determined as a function of output level. A transfer characteristic with inductive load is shown as curve B in Fig. 2. From a comparison of the slopes of the two curves of Fig. 2, the factor

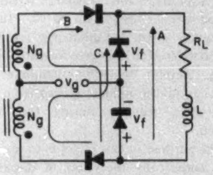


Fig. 4. Circuit showing load current paths

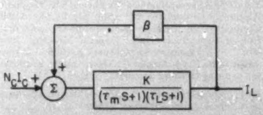


Fig. 5. Representation of feedback caused by nonideal discharge rectifiers

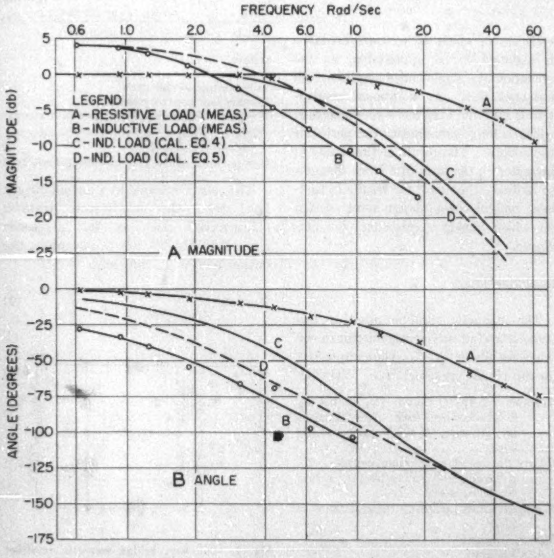


Fig. 3. Frequency response characteristics

Vol. 7
Pt. 1
1956

$(1-\beta K')$ is found at an output level of 0.2 amp to be

$$(1-\beta K')|_{I_L=0.2}=0.53 \qquad (6)$$

The frequency response for the circuit with the inductive load indicated in Fig. 1, for a small signal at an output level of 0.2 amp, is shown as curves B in Fig. 3. Two predictions are made for comparison. Curves C in Fig. 3 are predicted with the use of the approximation of perfect discharge rectifiers given in equation 4. Curves D are predicted with the use of the transfer function of equation 5 with the value of $(1-\beta K')$ found from the steady-state transfer characteristics and given by equation 6.

The results presented in Fig. 3 indicate that the addition of the positive feedback loop does not give absolutely accurate prediction of the experimental results; however, the accuracy indicates that such a feedback probably does occur but not wholly in the simple manner depicted. Nevertheless, the addition of the positive feedback loop does give results which can be useful in design work, especially with analogue computers.

Conclusions

The effect of nonideal discharge rectifiers on the dynamic operation of a full-wave magnetic amplifier with inductive load has been shown to be empirically describable as a nonlinear positive feedback. More detailed study is necessary to determine exactly how the feedback factor β varies with all the circuit parameters.

The problem of nonideal discharge rectifiers arises in connection with the application of selenium rectifiers. When silicon-junction rectifiers are used, this problem essentially disappears. However, in some cases, additional resistance be place in series with the discharge rectifiers to acquire special characteristics obtainable with the resulting nonlinear feedback.

The concept of a positive feedback to account for imperfect discharge rectifiers has been used to obtain an improved design for an aircraft alternator regulator. this concept proved especially useful in simplifying computer studies of the complete system.

References

1. MAGNETIC AMPLIFIERS WITH INDUCTIVE D-C LOAD, H. F. Storm. *AIEE Transactions*, vol. 73, pt. I, Nov. 1954, pp. 466–75.

2. THE OPERATION OF MAGNETIC AMPLIFIERS WITH VARIOUS TYPES OF LOADS, PART II, L. A. Finzi, R. R. Jackson. *Ibid.*, July 1954, pp. 279–88.

3. INDUCTIVE LOAD INSTABILITY IN MAGNETIC AMPLIFIERS, H. I. Leon, A. B. Rosenstein. *Ibid.*, vol. 74, pt. I, March 1955, pp. 8–20.

4. RESPONSE TIME OF MAGNETIC AMPLIFIERS, E. L. Harder, W. F. Horton. *Ibid.*, vol. 69, 1950, pp. 1130–38.

Standing-Wave Ratio of Inaccessible Load

CHARLES POLK
ASSOCIATE MEMBER AIEE

Synopsis. Formulas are given for the minimum and maximum possible voltage standing-wave ratio (VSWR) of an inaccessible load fed by an imperfect transmission line. The formulas involve measurements at the accessible input terminals of the transmission line. Required data are the VSWR existing when the line is terminated by a known, perfectly matched load and the VSWR measured when the line is terminated by the unknown load.

WHEN a determination of the input VSWR of an antenna located on a high tower is required, it is frequently difficult to perform measurements directly at the antenna terminals. It is, however, possible to disconnect the antenna from the transmission line and to replace it, temporarily, by a load matched exactly to the nominal characteristic impedance of the line. Measurements can then be performed on the ground, at the input terminals of the transmission circuit.

It is thus possible to determine not only the VSWR (S) of the transmission line terminated by the antenna, but also the VSWR (S_1) of the transmission line terminated by a nominally matched load.

If S_1 is equal to one, the transmission circuit is perfect, or at least all reflections along the transmission line cancel each other at the input terminals. In this case one may assume that S is the actual VSWR of the antenna. Usually, however, $S_1 \neq 1$ indicating some imperfections along the transmission line which may be due to such causes as imperfect insulators, bends, joints, or transitions. The problem is then to deduce from S_1 and S the possible minimum and maximum values of the VSWR which would be measured directly at the antenna terminals.

While the present paper resulted from the antenna problem, the equations given are, of course, applicable to any analogous situation. The antenna may be replaced by any unknown load and the transmission circuit may consist, for example, of coaxial lines or wave guides. The frequency involved is assumed to be high enough for VSWR data to be of interest.

Attenuation is neglected in the present analysis and therefore the equations are strictly valid only if the transmission circuit has no dissipation loss. However, if the VSWR's of the system are small, small losses in the line will not account for any appreciable difference between the VSWR's measured at opposite ends of the transmission lines. For instance, compare reference 1, page 148: If the transmission-line dissipation loss is $^1/_2$ decibel (89-per-cent line efficiency), a 1.5 VSWR at the load termination will appear as a 1.432 VSWR at the input end.

Equations

Let

S = VSWR measured at the input terminals of the transmission circuit when the unknown load is connected to the far end

S_1 = VSWR measured at the input terminals of the transmission circuit when a perfectly matched load is connected to the far end. A perfectly matched load is one matched exactly to the nominal characteristic impedance of the transmission line at the measurement frequency

$S_{2\max}$ = maximum possible VSWR of the unknown load

$S_{2\min}$ = minimum possible VSWR of the unknown load

Paper 56-32, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted July 18, 1955; made available for printing November 15, 1955.

CHARLES POLK is with the University of Pennsylvania, Philadelphia, Pa.

The author wishes to express his appreciation to his colleague, Dr. P. P. Lombardini, for helpful discussions on the subject of this paper.

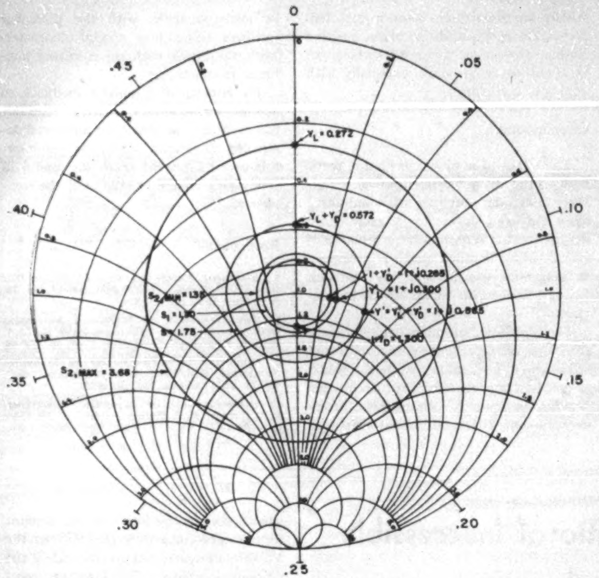


Fig. 1. Illustration of numerical example

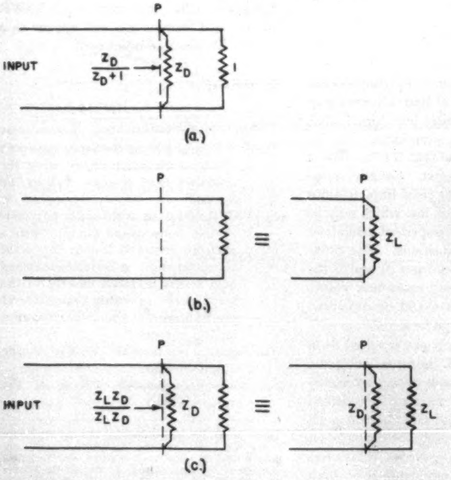


Fig. 2. Transmission-line circuit considered for the derivations of equations 1 and 2

When the conductor and dielectric losses in the transmission circuit are zero, the relations for $S_{2\max}$ and $S_{2\min}$ are

$$S_{2\max} = \frac{S}{S - SS_1 + 1} \quad SS_1 \leq 1 + S \qquad (1)$$

$$S_{2\min} = \frac{S}{S_1} \quad S > S_1$$

$$S_{2\min} = \frac{S_1}{S} \quad S < S_1 \qquad (2)$$

Equation 1 is exact while equation 2 is valid only for VSWR's less than about three. The nature of the approximation is indicated in the derivation.

Numerical Example

Standing-wave ratio measured at input terminals when the unknown load is connected to the transmission line, $S = 1.75$. Standing-wave ratio measured at input terminals when the transmission line is terminated by a perfectly matched load, $S_1 = 1.30$.

From equation 1

$S_{2\max} = 3.68$

From equation 2

$S_{2\min} = 1.345$

These results may be checked approximately by reference to the Smith-chart, Fig. 1. For convenience, admittances are considered on the Smith-chart rather than impedances.[2]

It is shown in the subsequent derivation that $S_{2\max}$ corresponds to real values of $Y_D = 1/Z_D$ and $Y_L = 1/Z_L$. Note that the matched load in parallel with the discontinuity gave a VSWR equal to 1.30, therefore $Y_D = 0.30$. The VSWR $S_{2\max} = 3.68$ found by equation 1 corresponds to a real admittance Y_L equal to 0.272. The sum of Y_L and Y_D is equal to 0.572 which corresponds to the measured standing-wave ratio $S = 1.75$.

To check $S_{2\min}$ assume Y_D as purely reactive; corresponding to $S_1 = 1.30$ $Y_D' = \pm$ j0.265 is obtained. When this Y_D' is placed in parallel with a load admittance $Y_L' = 1 \pm$ j0.30, corresponding to $S_{2\min} = 1.345$, $Y = 1 \pm$ j0.565 is obtained. This admittance corresponds again to the measured standing-wave ratio $S = 1.75$.

Derivation of Equations

The complex reflection coefficient $\Gamma = |\Gamma| \exp j\delta$ is defined as the ratio of the reflected to the incident wave. It is well known that the value of Γ at a point P on a transmission line of characteristic impedance K is given by

$$\Gamma = \frac{1-Z}{1+Z} \qquad (3)$$

where $Z = Z_T/K$ is the normalized value of the load impedance Z_T presented to the incident wave.[3]

In the following derivation all impedances are normalized. The impedance Z_D at point P in Fig. 2 represents the effective transmission-line discontinuity. If a matched load ($Z_T = K$) is connected to the line, the total input impedance at P becomes $Z_D/(Z_D+1)$ and, from equation 3, the complex reflection coefficient Γ_1 at point P, corresponding to this load, is given by

$$\Gamma_1 = \frac{1}{2Z_D+1} \tag{4}$$

Consider next the transmission line of Fig. 2 without the discontinuity ($Z_D=\infty$), but terminated by an unknown impedance, Fig. 2(b). The normalized value of this impedance, referred to point P, is Z_L. For this case the reflection coefficient at P is

$$\Gamma_1 = \frac{1-Z_L}{1+Z_L} \tag{5}$$

Finally, consider the line with the discontinuity and the unknown load impedance, Fig. 2(c). The reflection coefficient at P is now

$$\Gamma = \frac{Z_L+Z_D-Z_LZ_D}{Z_L+Z_D+Z_LZ_D} \tag{6}$$

Expressing Z_L in terms of Γ and Γ_1 by use of equations 4 and 6 and by substitution into equation 5, the following is obtained:

$$\Gamma_2 = \frac{\Gamma-\Gamma_1}{\Gamma\Gamma_1-2\Gamma_1+1} \tag{7}$$

The reflection coefficients Γ, Γ_1, and Γ_2 are complex numbers; thus, one may write

$$\Gamma_2 = \frac{|\Gamma|e^{j\alpha}-|\Gamma_1|e^{j\beta}}{|\Gamma\Gamma_1|e^{j(\alpha+\beta)}-2|\Gamma_1|e^{j\alpha}+1} \tag{8}$$

The square of the absolute value of Γ_2 is given by

$$|\Gamma_2|^2 = \frac{|\Gamma|^2+|\Gamma_1|^2-2|\Gamma\Gamma_1|\cos(\alpha-\beta)}{|\Gamma\Gamma_1|^2+4|\Gamma_1|^2+1-4|\Gamma||\Gamma_1|^2\cos\alpha - 4|\Gamma_1|\cos\beta+2|\Gamma\Gamma_1|\cos(\alpha+\beta)} \tag{9}$$

By taking the partial derivatives with respect to α and β and equating to zero, it is found that one of the following relations must be satisfied if Γ_2 is to assume an extreme value:

$$\begin{aligned} &\alpha=0, \quad \beta=0 \\ \text{or } &\alpha=\pi, \quad \beta=0 \\ \text{or } &\alpha=0, \quad \beta=\pi \\ \text{or } &\alpha=\pi, \quad \beta=\pi \end{aligned} \tag{10}$$

Substitution back into equation 9 gives:

$$\alpha=\pi,\ \beta=0 \quad |\Gamma_2|_{\max} = \left|\frac{|\Gamma|+|\Gamma_1|}{1-|\Gamma\Gamma_1|-2|\Gamma_1|}\right| \tag{11}$$

$$\alpha=\pi,\ \beta=\pi \quad |\Gamma_2|_{\min} = \left|\frac{|\Gamma_1|-|\Gamma|}{1+|\Gamma\Gamma_1|+2|\Gamma_1|}\right| \tag{12}$$

However, examination of equation 4 shows that β equal to π would require Z_D to be a real negative number. This is impossible if Z_D is a passive element. When $|\Gamma|<0.5$ and $|\Gamma_1|<0.5$, a physically realizable low value of $|\Gamma_2|$ is given by assuming $\alpha=\beta$ and $\cos\beta=|\Gamma_1|$. Substitution of these values of α and $\cos\beta$ into equation 9 leads to

$$|\Gamma_2|_{\min} \approx \left|\frac{|\Gamma_1|-|\Gamma|}{1-|\Gamma\Gamma_1|}\right| \tag{13}$$

To obtain the VSWR's corresponding to $|\Gamma_2|_{\max}$, equation 11, and $|\Gamma_2|_{\min}$, equation 13, note that

$$|\Gamma| = \frac{S-1}{S+1} \quad |\Gamma_1| = \frac{S_1-1}{S_1+1} \quad |\Gamma_2| = \frac{S_2-1}{S_2+1} \tag{14}$$

Substitution of equation 14 into equations 11 and 13 leads to equations 1 and 2.

References

1. ELECTRONIC MEASUREMENTS (book), F. E. Terman, J. M. Pettit. McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1952.
2. TRANSMISSION LINE CALCULATOR, P. H. Smith. *Electronics*, New York, N. Y., vol. 12, Jan. 1939, p. 29.
3. AN IMPROVED TRANSMISSION-LINE CALCULATOR, P. H. Smith. *Ibid.*, vol. 17, Jan. 1944, p. 130.
4. ELECTROMAGNETIC WAVES (book), S. A. Schelkunoff. D. Van Nostrand Company, Inc., New York, N. Y., 1943 (fifth printing, 1947).

Simplified Printing Telegraph Switching and Integrated Data Processing

J. B. BOOTH
ASSOCIATE MEMBER AIEE

R. H. KLICH
NONMEMBER AIEE

Synopsis: Receiving and transmitting printing telegraph equipment, whether used in the simplest or the most complex circuit, has generally required some form of auxiliary equipment for switching and remote control. In many cases this auxiliary equipment is complex, costly, and requires as much space and electric power as the controlled equipment. A new concept in printing telegraph switching equipment has been realized in the development of the Teletype model *28* typing unit which may be utilized to perform many electrical and mechanical functions for selective calling, remote control, integrated data processing, and other special functions.

THE MODEL *28* typing unit,[1] shown in Fig. 1, is a completely new instrument designed primarily for high-speed printing telegraphy. The unit is small in size, light in weight, and at a given speed of operational requires considerably less maintenance than normally recommended for apparatus of this type. This unit operates with greater ease and efficiency at 100 words per minute than conventional units at slower speeds.

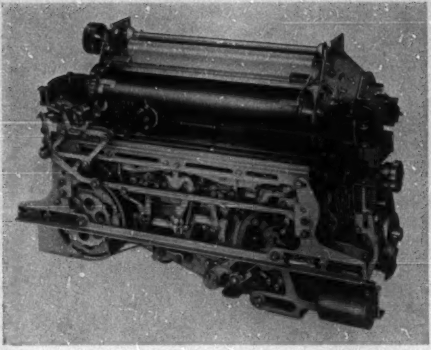
Every typing unit is equipped with a stunt box which controls the common functions, such as shifting the type box from lower to upper case, returning the printing carriage to the left-hand margin, line feeding, operating the signal bell, etc. In addition, the stunt box, as illustrated in Fig. 2, is capable of performing many electrical and mechanical functions.

The typing unit selector responds to a start-stop 5-unit (Baudot) signaling code consisting of current and no-current intervals, referred to as marking and spacing impulses, respectively. The selector mechanism of the typing unit converts these electric signals into mechanical motions which are transferred to a set of code bars to control the position of the type box for printing the received character, and to determine the function performed by the stunt box.

The stunt box, located directly behind the code bar assembly, has 42 slots, each of which will accommodate a function bar and its associated mechanism. These function bar mechanisms can be employed, in conjunction with the code bars, to initiate a mechanical or electrical action within the machine or at a remote location. Fig. 3 is a rear view of the typing unit with the stunt box removed to expose the code bar assembly.

Paper 56-28, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted November 2, 1955; made available for printing November 14, 1955.

J. B. BOOTH and R. H. KLICH are with the Teletype Corporation, Chicago, Ill.



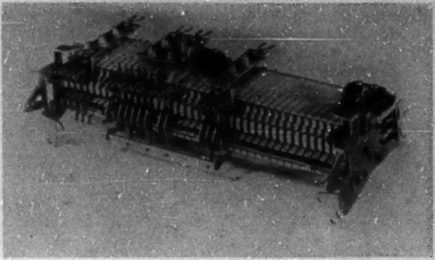


Fig. 1 (left). The model 28 typing unit

Fig. 2 (above). The typing unit stunt box

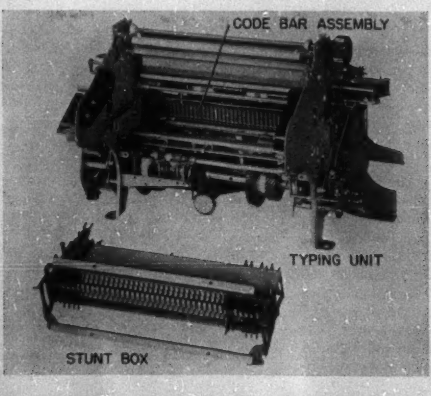
The code bars are moved to the left or right in response to marking or spacing elements of the Baudot code. A series of square notches extend across the rear of each code bar. The front of each function bar contains a series of tines which are formed to the left or right according to the marking or spacing pattern of the code combination to which the bar is to respond. The function bar-code bar relationship is illustrated in Fig. 4. The operation of any function bar mechanism is dependent upon the position of the code bars when sensed by the function bars. If the position of the code bar notches and function bar tines match, the function bar will enter the recessed portion of the code bar and condition the associated function bar mechanism for operation. The function bar, following its selection, is moved towards the rear of the typing unit performing a function dependent upon the arrangement of its associated mechanism. To change the code combination of any position in the stunt box only requires changing the function bar. The stunt box is readily removable to facilitate this change.

A simplified approach to many of the problems of switching and control is now possible using the model *28* typing unit. Usage in such applications as selectively calling stations in telegraph networks and integrated data processing are described.

Selective Calling of Model 28 Page Printer Sets

Many teletypewriter networks are equipped to permit stations selectively to call one another. This operation, referred to as selective calling, may be accomplished in many ways. Some systems utilize discrete telegraph characters to select the desired stations, some employ pulse duration or pulse coincidence signal coding, while others use frequency or tone keying. Regardless of the method used, the selection of teletypewriter stations has generally required some form of auxiliary control apparatus. The model *28* selective calling typing unit utilizes discrete telegraph characters, and requires no auxiliary control apparatus.

A review of the present usage of the Baudot code, as related to teletypewriters, will aid in understanding the operation and potential of the model *28* selective calling typing unit. The Baudot code contains five elements, each of which may assume a marking or spacing condition. This arrangement provides 32 distinct code combinations. All teletypewriters, however, are capable of performing approximately twice as many operations (printing or performing a function) as the number of available code combinations. This is accomplished by reserving two of the original code combinations to shift and unshift the platen (or type box in the case of the model *28*), placing the unit in the upper or lower case as desired.



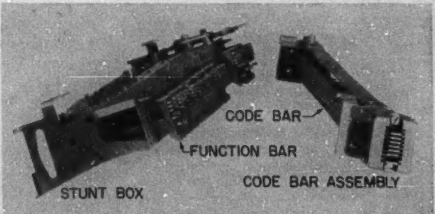


Fig. 3 (left). Rear view of the typing unit with the stunt box removed

Fig. 4 (above). Stunt box and code bar assembly showing relationship of function and code bars

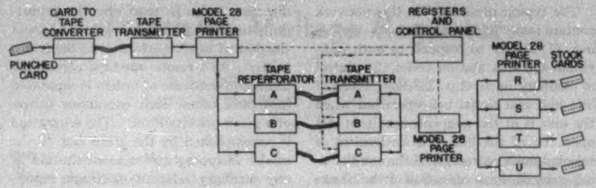


Fig. 5. Block diagram of integrated data processing system, utilizing model 28 equipment

The model *28* typing unit, when arranged for selective calling, is equipped with a device which activates or suppresses the printing mechanism. This arrangement permits an expanded use of the 32 code combinations of the Baudot code beyond that made available by the shift and unshift mechanism, since the unit can now be placed in a nonprint condition. In this condition the selector and stunt box are active, but the printing mechanism is suppressed; consequently, the same code combination that resulted in the printing of two different characters (one in the lower and one in the upper case) may be used to perform two additional functions, either electrical or mechanical, without the typing unit printing. These additional functions are prevented from occurring when the unit is in the print condition by positioning an auxiliary code bar to block all function bars arranged to perform special functions.

Suppression of the printing mechanism is accomplished by blocking the type box clutch. This blocking mechanism is controlled by elements in the stunt box which are set up to respond to a particular character or character sequence. Reception of this character or character sequence causes the blocking mechanism to unblock the type box clutch, placing the unit in the print condition. In this condition, the typing unit will operate as a normal teletypewriter. To return the unit to the nonprint condition a character sequence differing from that which placed the unit in the print condition causes the blocking mechanism to block the type box clutch. Code combinations selected for this latter sequence are preferably those which will result in no printing.

The versatility of the model *28* typing unit equipped with the blocking mechanism permits many variations of selective calling. The stunt box flexibility, as applied to selective calling, might best be illustrated by a brief description of two systems which are currently being placed in service.

BELL SYSTEM INSTALLATIONS

Teletype model *28* page printer equipment, now being placed in selective calling service by the Bell System, may be arranged to satisfy many diversified customer requirements. In their installations, each unit will shift to the print condition upon the reception of its preassigned character or sequence of characters, referred to as call directing characters (CDC). Units may be equipped to respond to more than one call directing character or character sequence so that a group of units, or all units, on a circuit may be placed in the print condition by the transmision of a single CDC. In addition, all units respond to two common codes, an end-of-address and an end-of-message code; the first shifts all units to a nonselect condition, and the second places the units in a nonprint and select condition.

As previously described, the position of the blocking mechanism determines whether the unit is in the print or nonprint condition. This mechanism is controlled by the call-directing and end-of-message codes. The units will be in the select or nonselect condition depending upon the position of an auxiliary (lockout) code bar. This code bar is controlled by elements in the stunt box which are set up to respond to the end-of-address and end-of-message codes. Reception of the end-of-address code causes the auxiliary code bar to be positioned to prevent selection of CDC function bars. Reception of the end of message code causes the auxiliary code bar to be positioned so as to allow selection of CDC function bars.

Any station desiring to communicate with one or more stations on the circuit may select them by transmitting his own and their assigned call directing characters. The end-of-address code sequence follows the transmission of the last CDC. All stations whether selected or not will have further selection suppressed, so that they will not respond to any selective calling codes except the end of message code. Call directing characters may, therefore, be used indiscriminately in the message without causing false selections. All stations not called will have printing suppressed and, therefore, will receive none of the communication. At the conclusion of the communication the end of message code sequence is transmitted, placing all units in the nonprint-select condition. In this condition, printing is suppressed and the stations are again capable of responding to call directing characters. The end-of-address code may be any character or character sequence. The end-of-message code, however, must be a sequence containing a minimum of two characters to insure that it will not appear in the body of the communication.

The number of stations possible in a network depends upon the character sequence used for call direction. If the 26 code combinations of the alphabet are used, a single call-directing character will provide 26 different station calls. If a 2-character sequence is employed, a total of 676 different sequences are possible. A 3-character sequence will provide 17,576 different call-directing sequences.

The selective calling stunt box arranged for Bell System Service also performs the following auxiliary functions: (1) The first character received by each unit, following the end-of-message code sequence, causes a busy lamp to light on all units on the circuit; this lamp will flash each time a space character is transmitted whether the unit has been selected or not. (2) The copy lights of each unit are turned on only when the unit is selected. (3) The keyboard lock mechanism of the units not selected is operated upon receipt of the end-of-address code to prevent inadvertent operation of the keyboard. (4) The line feed and signal bell operations are suppressed when the unit is not selected (nonprint condition).

Many optional features may be provided to meet particular customer requirements. These include the control of tape transmitters and reperforators at remote locations; motor control; the equipment of units to serve as spares in an installation containing many units, each having a different CDC; and the provision of a divide to shift the unit manually into the print condition when line monitoring is desired.

CAA INSTALLATION

The Civil Aeronautics Administration (CAA) is installing a nation-wide selective-calling flight-control network, utilizing the Teletype model *28* page printer set. This network consists of approximately 400 stations divided into 24 circuits. Each circuit contains from 10 to 30 stations, including relay stations for transmission to and from adjacent circuits. An expressway or trunk line is provided for transmission from coast to coast. The CAA installation and operating procedure differ from those of the Bell System. However, the model *28* page printer sets are essentially the same, differing primarily in the stunt box arrangements.

The CAA precedes the call directing characters with a 3-character code sequence that conditions the printer to respond to a station call. Upon receipt of the conditioning code a mechanism in the stunt box positions the auxiliary code bar (lockout code bar in the Bell System units) to allow the station call function bars to be selected. This code bar is held in the operated position for only the succeeding character. Consequently, if the first character of the station call does not follow the last character of the conditioning code the auxiliary code bar will return to its unoperated position, blocking the station call function bars associated with the blocking mechanism. The combination of the conditioning code and station code is a sequence of six characters which would not normally appear in the message. Use of the conditioning code permits the selection of any station at any time during a communiction without dropping out stations already selected.

Any station may communicate with one or more stations on the same circuit by transmitting the conditioning code and the desired station code. A selected unit is placed into the print condition upon receipt of the last character of the station code. An end-of-message sequence, similar to that employed in the Bell System installations, concludes the communication, placing all units in the nonprint condition.

To communicate with stations in adjacent areas requires the selection of a relay station in the circuit of the originating station. The relay stations contain reperforator-transmitter distributors controlled by model *28* page printer sets. Both of the units are placed into the circuit upon the receipt of the conditioning code and relay station code. The message is reproduced in the form of perforated tape by the reperforator portion of the reperforator-transmitter distributor. At the same instant, the circuit in the adjacent area is sampled to determine if it is busy. If the circuit is available, the tape is read by the transmitting portion of the reperforator-transmitter distributor and retransmitted. If the circuit is busy, the message is stored in the perforated tape for retransmission when the circuit becomes available.

The typing units used in this network contain many additional features, including: (1) ability to respond to both individual and area station codes, (2) control of auxiliary apparatus, (3) suppression of line feed and signal bell operation when the unit is in the nonprint condition, (4) operation of the line feed mechanism of selected units on receipt of the conditioning code, (5) local operation of the blocking mechanism for line monitoring, and (6) the ability of associated transmitter-distributor to shift the typing unit to the print condition without transmitting the conditioning or station codes.

Integrated Data Processing With Teletypewriters

Integrated data processing, as the name implies, is the integration data processes to achieve an automatic flow of information from one data processing machine to another. Mechanical and electronic data-processing devices of many types are available to perform the tedious tasks that are a part of office routines. Integration of this data-processing equipment may be simplified by the use of the Teletype model *28* page printer set in combination with perforated-tape-handling equipment. This simplification is possible because of the ability of the typing unit to perform many switching, sorting, and checking functions besides recording desired data.

An explanation of an order handling system providing rapid service on parts orders will illustrate the potential use of model *28* equipment in this field. This system accepts random orders for separation in accordance with priority levels, and then directs the items of each order to the proper stock area.

The input to the system, as shown in Fig. 5, is a punched card prepared by a key punch operator for each item on the order. The cards contain the necessary parts ordering information and the required directing codes which indicate the priority of the order and the item stock location.

The punched cards are placed into a card-to-tape converter which prepares a continuous common-language 5-unit (Baudot code) perforated tape. In addition to converting the information from card to tape, the converter inserts additional codes which are required for conditioning and releasing the switching equipment.

The prepared tape is placed into a transmitter of the type used for normal telegraph transmission. The information contained in the tape is transmitted to a model *28* page printer set. The order directing codes is "read" by the typing-unit stunt box mechanims which connect the desired tape reperforator to the transmitter. As a result, one tape containing a random sequence of orders is separated into three tapes, each containing orders of a given priority level. The separation is accomplished by the stunt box of the model *28* typing unit without the aid of any auxiliary relay or electronic equipment.

In addition to performing switching functions, the typing unit is available for recording the transmitted information. An invoice, shipping paper, or inventory record may, therefore, be prepared at the same instant and by the machine performing the circuit switching functions. Registers on a control panel are operated by contacts on the stunt box to provide a visible indication of the number of items, according to priority level and stock location, to be dispersed to stock areas.

The tapes prepared by the three reperforators *A*, *B*, and *C* are placed into their respective transmitters *A*, *B*, and *C* for transmission to the desired stock areas; the three tape transmitters operate one at a time under the control of a dispatcher. The stock area directing codes are "read" by a second model *28* typing unit which connects the desired stock area typing unit to the transmitter then operating. The stock area printer sets produce a stock ticket containing, in printed form, the same ordering information that appeared in the original punched cards. The typing unit performing this sorting operation operates registers to record the number and priority level of items dispersed to the stock areas.

The registers, operated by the two typing units performing the switching, provide the dispatcher with the load on each stock area, both in the number of items waiting for distribution and the number of items dispersed. Utilizing this information, the dispatcher is able to move the high priority orders first, but without overloading any particular stock area.

Conclusion

The experience in developing equipment for the systems described, together with the proved versatility and reliability of the stunt box, is expected to lead to an expanded use of the Teletype model *28* page printer set in selective calling, integrated data processing, and automation.

Reference

1. A Step Forward in Printing Telegraphy, A. S. Benjamin, W. J. Zenner. *AIEE Transactions*, vol. 73, March 1954, pp. 10–15.

Vol.
Pt. 1
1956

Transistor Voltage Regulator

R. H. SPENCER
ASSOCIATE MEMBER AIEE

T. S. GRAY
FELLOW AIEE

THIS paper describes a transistor voltage regulator circuit which takes advantage of the high voltage gain and low emitter resistance exhibited by the junction transistor. Emphasis is on a regulator circuit using a single transistor.

Circuit

The circuit of the basic transistor voltage regulator is illustrated in Fig. 1. Like other voltage regulators, this transistor regulator may be analyzed in terms of a 2-terminal-pair network as shown in Fig. 2.

The quantity to be regulated is the output voltage V_o which is a function of the two independent variables, the input voltage, V_{in} and the output current I_o, thus:

$$V_o=F(V_{in},I_o),$$

Variations of this output voltage, which are of concern here, may, when infinitesimal, be expressed in derivative form:

$$dV_o=\frac{\delta V_o}{\delta V_{in}}\,dV_{in}+\frac{\delta V_o}{\delta I_o}\,dI_o$$

For variations of noninfinitesimal size, the following expression is approximately correct provided the circuit is operated in the essentially linear region:

$$\Delta V_o=\frac{\delta V_o}{\delta V_{in}}\,\Delta V_{in}+\frac{\delta V_o}{\delta I_o}\,\Delta I_o$$

For this general regulator, the factor $\delta V_o/\delta V_{in}$ is called the stabilization factor F, and $-\delta V_o/\delta I_o$ is referred to as the output resistance, R_o. In terms of these definitions, the expression may be rewritten as follows:

$$\Delta V_o=F\Delta V_{in}-R_o\Delta I_o$$

The two factors F and R_o may be used to determine the actual stabilization for any output load resistance.

To determine the two describing parameters F and R_o for the transistor regulator circuit shown in Fig. 1, the incremental equivalent circuit of the regulator shown in Fig. 3 may be used.[1] Since F is $\Delta V_o/\Delta V_{in}$ when ΔI_o is made zero, the equivalent circuit for determination of F becomes that shown in Fig. 4. For this circuit $\Delta V_o/\Delta V_{in}$, and therefore F, is $r_b/(r_b+r_c)$.

To evaluate the output resistance R_o, which is $\Delta V_o/\Delta I_o$ when ΔV_{in} is made zero, the equivalent circuit of Fig. 5 is used. Loop equations for the voltages in this circuit are

$$0=\Delta I_c(r_c+r_b)+\Delta I_e(r_b+\alpha r_c)$$

$$\Delta V_o=\Delta I_c(r_b)+\Delta I_e(r_e+r_b)$$

and ΔI_e is thus:

$$\Delta I_e=\frac{\Delta V_o(r_c+r_b)}{(r_c+r_b)(r_e+r_b)-(r_b+\alpha r_c)(r_b)}$$

With the denominator rewritten the expression becomes:

$$\Delta I_e=\frac{\Delta V_o(r_c+r_b)}{r_c\left[r_e\left(1+\frac{r_b}{r_c}\right)+r_b(1-\alpha)\right]}$$

Since for most junction transistors $r_b/r_c<<1$, the output resistance is very nearly:

$$R_o=r_e+r_b(1-\alpha)$$

As the frequency of the output variation increases, the value of R_o increases until it reaches r_e+r_b at frequencies well above the alpha cutoff frequency.

For a representative transistor, the Raytheon *CK721*, the values of the parameters are

$r_c=700$ *K* ohms
$r_b=0.35$ *K* ohms
$r_e=0.020$ *K* ohms
$\alpha=0.975$

Using this transistor, one obtains for F, the stabilization factor:

$$F=\frac{r_b}{r_b+r_c}=\frac{0.35}{700}=\frac{1}{2{,}000}$$

The output resistance $R_o=r_e+r_b(1-\alpha)$ $=29$ ohms.

Thus, for low-voltage low-current loads, this circuit acts as an excellent regulator.

Voltage References

Silicon junction diodes now available serve as good voltage reference elements because in the so-called Zener breakdown region the voltage is relatively constant over a wide range of currents. Fig. 6 illustrates this Zener region as it appears in a representative silicon junction diode.

The voltage at which the so-called Zener breakdown occurs is fairly stable, exhibiting a temperature coefficient between −0.04 and +0.010 per degree centigrade.[2] Where smaller variations with temperature are required, the use of back-to-back Zener diodes allows an order of magnitude improvement in temperature coefficient.[2] This improvement occurs because the temperature coefficient in the Zener region can be made to have the opposite sign as that in the forward region, causing a first-order canceling of temperature coefficient.

The dynamic resistance of the Zener region is relatively low, ranging between 10 and 1,000 ohms. About 100 milliwatts of power can be dissipated in the Zener region by available junction diodes. Diodes of this type can be constructed with breakdown voltages ranging from a few volts to several hundred volts,[2] indeed a desirable feature of a voltage-reference unit.

In the transistor regulator shown in Fig 1, the reference element must pass a small fraction of the load current. Thus, the reference element must operate over a current range approximately $(1-\alpha)$ times the range of load currents. For the above 5-milliampere regulator, the biasing current of the reference must be at least 5(0.025) or 0.125 milliampere. In practice, the bias should be larger than this minimum value to avoid operation in the curved, low-current region of the Zener characteristic.

It should be noted that the incremental resistance of the reference element R_{ref} appears in the equivalent circuit as an increase in the base resistance of the transistor. This increase in equivalent base resistance will appear in the regulator output resistance as follows:

$$R_o=r_e+R_b(1-\alpha)$$

where

$$R_b=r_b+R_{ref}$$

Representative Circuits

Fig. 7 shows a practical voltage regulator circuit which takes into account the preceding principles. Modifications and extensions of this basic circuit are readily

---

Paper 56-41, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 19, 1955; made available for printing November 17, 1955.

R. H. SPENCER and T. S. GRAY are with the Massachusetts Institute of Technology, Cambridge, Mass.

The research reported in this paper was supported by the Office of Naval Research and the Navy Bureau of Ships.

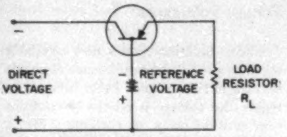

DIRECT VOLTAGE
REFERENCE VOLTAGE
LOAD RESISTOR R_L
−
+

Fig. 1. Basic transistor voltage regulator

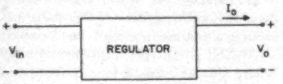

I_o
V_{in}
REGULATOR
V_o
+
−

Fig. 2. Generalized voltage regulator

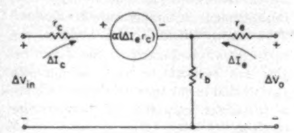

r_c
$\alpha(\Delta I_e r_c)$
r_e
ΔI_c
ΔI_e
r_b
ΔV_{in}
ΔV_o
+
−

Fig. 3. Incremental equivalent circuit of regulator

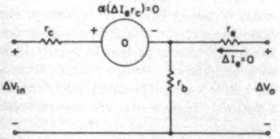

$\alpha(\Delta I_e r_c)=0$
r_c
0
r_e
$\Delta I_e=0$
r_b
ΔV_{in}
ΔV_o
+
−

Fig. 4. Circuit for determination of stabilization factor

made. Two transistors in cascade provide lower output resistance and better stabilization than the basic regulator. As shown in Fig. 8, this modification adds little to the complexity of the circuit.

The use of silicon junction transistors such as the type *901* of the Texas Instrument Company makes possible a higher output current than the use of similar germanium transistors, and permits operation to temperatures over 100 degrees centigrade. The use of the *n-p-n* transistors in this configuration results, of course, in a positive instead of a negative supply.

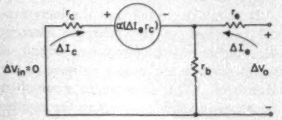

r_c
$\alpha(\Delta I_e r_c)$
r_e
ΔI_c
ΔI_e
$\Delta V_{in}=0$
r_b
ΔV_o
+
−

Fig. 5. Circuit for determination of output resistance

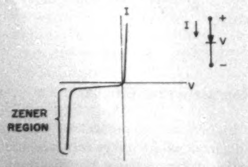


Fig. 6. Characteristic of silicon junction diode

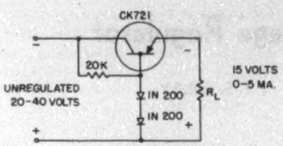

CK721
20K
UNREGULATED 20-40 VOLTS
IN 200
IN 200
R_L
15 VOLTS 0–5 MA.
−
+

Fig. 7. Circuit of a practical regulator

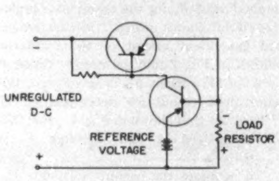

UNREGULATED D-C
REFERENCE VOLTAGE
LOAD RESISTOR
−
+

Fig. 8. Two-stage voltage regulator circuit

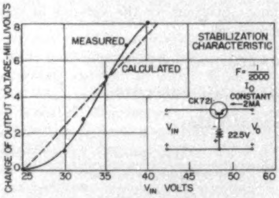


Fig. 9. Stabilization characteristic for regulator using a battery reference

Experimental Results

The performance of an experimental voltage regulator is shown in Figs. 9 and 10. The measured values of stabilization factor and output resistance agree well with those calculated. In this regulator, a battery is used as reference. Figs. 11 and 12 indicate the characteristics of a regulator using silicon junction diodes as the reference element. It will be noticed that the stabilization factor of this regulator is not as good as that of the regulator with a battery reference. This inferior stabilization factor results largely from the fact that the diode Zener voltage changes with the unregulated voltage much the same as the voltage of a voltage regulator tube changes with the unregulated voltage in a regulated power supply using a voltage regulator tube as the reference.

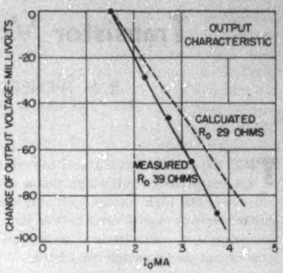


Fig. 10. Output characteristic for regulator using a battery reference

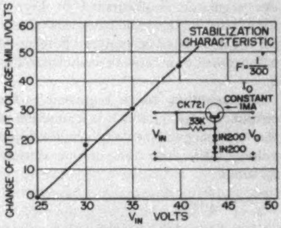


Fig. 11. Stabilization characteristic for regulator using a diode reference

Comparison With Vacuum-Tube Circuits

It is interesting to compare the transistor regulator and a vacuum-tube regulator which uses one tube. Only a pentode circuit can show as good a stabilization factor as does the transistor regulator. The pentode circuit suffers because the output resistance is $r_p/(\mu+1)$, which is seldom less than 200 ohms. Likewise, in a triode circuit, the output resistance is $r_p/(\mu+1)$, and the stabilization factor

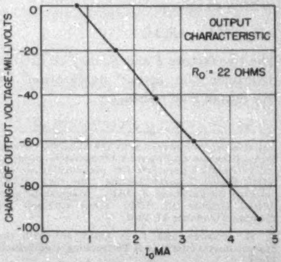


Fig. 12. Output characteristic for regulator using a diode reference

in a simple triode circuit is limited to $\mu/1$. Thus, the transistor circuit shows better inherent performance in these respects than single vacuum-tube circuits.

Conclusions

The simplicity of these circuits together with their good performance makes them well suited to regulate direct voltage within the limits of transistor capabilities.

References

1. Applied Electronics (book, second edition), T. S. Gray. John Wiley and Sons, Inc., New York, N. Y., 1954, chap. 13.

2. The Suitability of the Silicon Alloy Junction Diode as a Reference Standard in Regulated Metallic Rectifier Circuits, D. H. Smith. *AIEE Transactions*, vol. 74, pt. I, Jan. 1955, pp. 645–51.

3. Silicon P-N Junction Alloy Diodes, G. L. Pearson, B. Sawyer. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, no. 11, Nov. 1952, pp. 1348–51.

A Comparison of Three Common Emitter Transistor Servo Preamplifiers

A. N. DeSAUTELS
ASSOCIATE MEMBER AIEE

Synopsis: Three multistage transistor servo preamplifiers representing transformer coupling, resistance-capacitance (R/C) coupling, and direct coupling respectively are compared on the basis of design problems and performance. Experimental data referring to temperature stability and voltage response supplement the design and performance calculations for a particular application. Primary advantages and disadvantages of each type are presented, together with a brief discussion of application generalities.

ONE OF the more popular configurations used for transistor amplifiers is that of the grounded or common emitter, whose vacuum-tube counterpart is the grounded cathode triode. There are three standard types of multistage common emitter amplifiers, and these are classified in accordance with the form of coupling used. These three types are transformer-coupled amplifiers, R/C-coupled amplifiers, and direct-coupled amplifiers. To provide information which may be used as a basis of approach by the engineer becoming acquainted with the transistor application field, these types are compared in this paper from the viewpoint of design problems and performance.

Nomenclature

α = short-circuit current amplification factor
I_c = d-c collector current
I_{c_0} = d-c collector current for zero emitter current
V_c = d-c collector voltage
E_{bb} = d-c supply voltage
R_L = d-c load resistance
S = stability factor
e_{in} = a-c transistor input voltage
e_T = a-c transistor output voltage
G = power gain
r_b = equivalent transistor base resistance
r_c = equivalent transistor collector resistance
r_e = equivalent transistor emitter resistance
r_{in} = a-c transistor input impedance
r_0 = a-c transistor output impedance
r_L = a-c transistor load impedance
θ = voltage phase shift

Design Considerations

Preliminary Design Facts

Each amplifier covered consists of three stages of amplification using *PNP* alloy junction germanium transistors. The transistors in the first two stages were 50-milliwatt types with α approximately equal to 0.97. A Minneapolis-Honeywell *P6* 2-watt transistor was used in the third stage. The α and I_{c_0} values used in design calculations were selected from the manufacturer's data and test and test data as representative for the transistor types. The same load was used as the foundation of design for each type. Although a particular case is cited in this paper, the design principles and general comparison should, with few exceptions, hold true in most applications for which a choice of the three amplifier types can be made. Possible applications for amplifiers of the types considered could include use as a preamplifier driving a power transistor, vacuum tube, or magnetic amplifier output stage.

Performance is evaluated for one condition of load and source impedance. The load is a 50-ohm resistor coupled by a transformer presenting approximately a 4-kilohm impedance to the amplifier. Source impedance is about 10 ohms. Typical operating points were arbitrarily selected. In the experimental study, the same transistors were used in all three amplifiers. Calculations are based on a signal frequency of 400 cycles.

Design problems connected with certain applications may limit or dictate the type of coupling used. This study is based on the premise that any type of coupling is possible for the application being considered.

Transformer-Coupled Amplifier

Fig. 1 presents a 3-stage transformer-coupled amplifier. The interstate transformers used in transistor circuitry perform more efficiently at lower values of primary direct current. In addition, the d-c collector-to-emitter potential must be high enough to forestall clipping at higher signal levels. With these two points in mind, the first and second stages were designed for d-c operating points of V_c = 10 volts, I_c = 0.3 milliampere (ma) for a power supply of 28 volts. The third-stage operating point was arbitrarily selected as typical for the *P6* power transistor. Design of the d-c bias networks[1] was based on a stability factor S of 1.3. This stability factor is defined as the partial derivative of collector current I_c, with respect to zero emitter collector current I_{c_0}, or $S = dI_c/dI_{c_0}$.

Stage-by-stage design calculations are presented in the Appendix. The emitter capacitor serves as an a-c by-pass to reduce degeneration across the emitter d-c bias resistor.

R/C-Coupled Amplifier

A 3-stage R/C-coupled amplifier is shown in Fig. 2. Design of this circuit is straightforward. Since the need for lower values of collector current is not of as great concern with R/C-coupling, d-c operating points of V_c = 6 volts and I_c = 1 ma were selected for the first and second stages. If a typical operating point was assigned to the third stage however, a situation detrimental to per-

---

Paper 56-37, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 18, 1955; made available for printing November 16, 1955.

A. N. DeSautels is with the Minneapolis-Honeywell Regulator Company, Minneapolis, Minn.

formance would result when bias stabilization resistance values were calculated. The low value of the base bias resistors would have a shunting effect on the input to the third stage. This effect can be countered in two ways. One way would be to increase the power supply voltage for a given operating point. This would result in larger values for the bias resistances with a corresponding reduction in shunt effect and a related decrease in coupling network losses. An alternate way would be to select an operating point that would permit higher values of d-c bias resistance for a given power supply. Fig. 2 illustrates the second method. A stability factor *S* of 1.3 was used in the design of the d-c bias networks. The emitter by-pass capacitors should be as high as possible to minimize degeneration.

DIRECT-COUPLED AMPLIFIER

The interrelation of stages of the direct-coupled amplifier make necessary a special design approach. Fig. 3 presents a direct-coupled amplifier circuit designed for stability of performance with temperature variations. Stability can be obtained by fixing the base bias voltages and keeping the emitter current essentially constant. To accomplish this, it is desirable that the emitter circuit resistances be in the order of ten times greater than the collector circuit d-c resistances. In order words, $R_3 >> R_1$ and $R_4 >> R_2$.

The collector-to-ground direct voltage of one stage is the base bias voltage of the following stage. This value is approximately the same as the emitter voltage of that stage. These conditions introduce special problems where operating points and power supply voltage are concerned.

To maintain practical values of collector-to-emitter d-c voltage in the 3-stage amplifier treated here, the d-c power supply must be considerably greater than the 28-volt supply used with the transformer- and R/C-coupled amplifiers. Operation of a bias stabilized 3-stage direct-coupled amplifier with the emitter and collector circuit resistance relationships previously mentioned and with a 28-volt d-c supply would be inconceivable even at the low operating points utilized for the circuit of Fig. 3. For good bias stabilization, it is preferred that $R_3 >> R_1$ and $R_4 >> R_2$ by a factor of about 10. For the first stage, this means that the voltage across R_3 will be approximately ten times the voltage across R_1. At a d-c operating point of 2 volts, 1 ma, the voltage from *A* to ground using a 28-volt supply would be approximately 2.5 volts direct current, which would also be the approximate voltage from *B* to ground. Thus, from *B* to ground, 2.5 volts would be available to be dropped across T_2 and R_2. Obviously, this becomes impossible at the operating points shown and use of operating voltages lower than those shown would approach impracticality, particularly in the third stage.

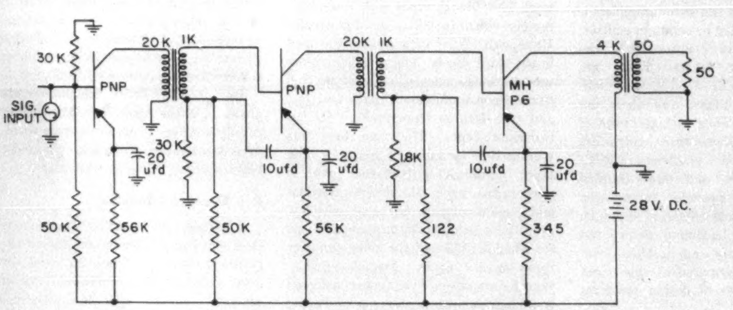
At $E_{bb}=100$ volts, it is possible to increase the operating voltages at the expense of some bias stabilization efficiency. Of course, bias stabilization effectiveness could be maintained at higher operating voltages than those shown by increasing the value of E_{bb}. D-c analysis of Fig. 3 reveals that the voltage from *A* to ground is 10 volts direct current, which is essentially the voltage from *B* to ground. $V_{c2}=3$ volts direct current, which leaves the voltage from *C* to ground and therefore *D* to ground approximately equal to 7 volts direct current. This permits sufficient voltage for the $V_{c3}=6$-volt drop and the direct voltage drop across the transformer primary.

Performance

POWER GAIN

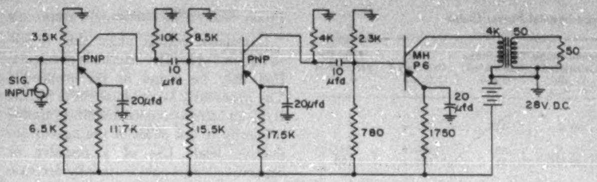
Because of the bilateral nature of the transistor (e.g., input impedance is a function of the load and output impedance is a function of the source), it is necessary that calculation of performance of a transistor amplifier be started from the last stage. Input impedance and power gains are calculated, and the input impedance combined with the coupling network becomes the load impedance for the preceding stage. The mechanics of performance calculations

Vol.
Pt. 1
195



	1ST & 2ND STAGES V_c = 10v d-c, I_c = 0.3ma d-c	3RD STAGE V_c = 25v d-c, I_c = 5ma d-c
r_c OHMS	$2(10^6)$	200K
r_b OHMS	1000	250
r_e OHMS	50	5
a	0.967	0.925

Fig. 1. Three-stage transformer-coupled amplifier



	1 ST STAGE	2 ND STAGE	3 RD STAGE
α	0.967	0.967	0.925
r_c MEGOHMS	0.5	0.5	0.3
r_b OHMS	400	400	250
r_e OHMS	20	20	5
V_c vd-c	6	6	20
I_c ma	1	1	4
I_{co} ma	10	10	300
R_L KILOHMS	10	4	.250
E_{bb} vd-c	28_v	28_v	28_v
S	1.3	1.3	1.3

Fig. 2. Three-stage R/C-coupled amplifier

for the amplifiers treated in this paper are presented in the Appendix.

Table I presents the comparative theoretical and experimental power gains for all three amplifiers. The coupling network losses add up to 3 decibels (db) for the R/C-coupled amplifier and about 1 db for the direct-coupled circuit. Increasing E_{bb} to 100 volts for the R/C-coupled network for the same operating points with related changes in the d-c bias resistances would reduce coupling network losses to 1 db with a resultant total calculated power gain for the specified load of about 83.5 db.

The transistor parameter values used in the calculations were obtained from average-characteristics curves for the operating points concerned and are necessarily approximate. The difference between calculated and experimental values could be accounted for by differences between approximate and actual parameter values of the transistors used.

It can be noted that changing the power supply voltage for the direct-coupled amplifier would have no effect on power gain if operating points are kept constant by varying the appropriate emitter circuit resistances. No change would be required in base bias resistance values.

Repeatability of power gain with transistor replacement was observed during interchanging, in the first two amplifier stages, ten transistors representing three manufacturers. About a 10-per-cent spread between highest and lowest values was noted for the over-all power gain. Feedback networks could be used to decrease this spread at the expense of some over-all power gain.

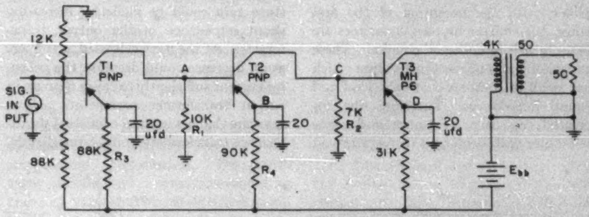
Input and Output Impedances

Table II presents a comparison of input and output impedances. Calculation processes are based on approximate parameter values for respective operating points and are detailed in the Appendix. The explanation for the comparatively high experimental input impedance may be that actual transistor parameters, particularly r_e and α are higher than those approximated for the calculations.

It would be well to emphasize at this point that, due to the probability expressed in the preceding statement, any comparison between theoretical and experimental data should be from a qualitative viewpoint.

Temperature Stability

Fig. 4(A) presents, for the three amplifiers under study, the comparison of power gain as a function of temperature increase. The stability of the R/C-coupled and direct-coupled amplifiers appears superior to that of the transformer-coupled amplifier. An increase in temperature has the effect of introducing changes in transistor parameters and direct current. Although these changes are minimized with bias stabilization, they still exist. For the transformers used in the transformer-coupled amplifier, small increases in the d-c levels will cause comparatively large decreases in the transformer shunt impedances. This means increased shunt impedance losses with an associated decrease in power transfer from transistor to load. Since a mismatch already exists between stages in R/C and direct coupling, the effect of small changes in transistor parameters is not as pronounced with these types of coupling.



	1 ST STAGE	2 ND STAGE	3 RD STAGE
α	0.967	0.967	0.925
r_c KILOHMS	400	450	250
r_b OHMS	300	330	200
r_e OHMS	15	15	5
V_c vd-c	2	3	6
I_c ma d-c	1	1	3
I_{co} ma d-c	0.010	0.010	0.300

Fig. 3. Three-stage direct-coupled amplifier

Table I. Calculated and Experimental Power Gains

Type of Amplifier		First-Stage Calculated G, db	Second-Stage Calculated G, db	Third-Stage Calculated G, db	Total Calculated G, db	Total Experimental G, db
Transformer-coupled	Each stage maximized	39.2	39.2	33	111.4	
	$R_{L\max} = 16$ kilohms	38.6	32.3	33	103.9	
	$R_L = 4$ kilohms	38.5	32.6	31	102.1	100
R/C-coupled	$R_{L\max} = 22.5$ kilohms	27.7	22.2	34.8	84.7	
	$R_L = 4$ kilohms	27.7	22.2	31.6	81.5	84
Direct-coupled	$R_{L\max} = 18.8$ kilohms	29.4	24.1	34.8	88.3	
	$R_L = 4$ kilohms	29.4	24.1	31.7	85.2	92

Table II. Comparison of Input and Output Impedances

Type of Amplifier	Calculated r_o, Kilohms	Calculated Third Stage r_{in}		Calculated Second Stage r_{in}		Calculated First Stage r_{in}		Experimental First Stage r_{in}
		$r_L{}^* = r_i$	$r_L = 4$ Kilohms	$r_L = r_o$	$r_L = 4$ Kilohms	$r_L = r_o$	$r_L = 4$ Kilohms	
Transformer-coupled	16.5	280	300	1,600	2,400	1,600	1,950	4,000
R/C-coupled	22.5	280	305	900	900	710	710	1,800
Direct-coupled	18.8	236	250	780	780	680	680	1,600

* r_L = output load on amplifier.

Fig. 4(B) shows power gain as a function of temperature decrease. Under these conditions, the capacitors in the circuit become a limiting factor. The characteristics of the capacitors are such that effective capacitive reactance and series resistance values increase with temperature decrease. Where the emitter by-pass capacitors are concerned this results in increased degeneration in the emitter circuits with a resultant gain decrease for R/C- and direct-coupled circuits. In the transformer-coupled amplifier, with the exception of the first stage, the emitter by-pass capacitors are entirely in the collector circuits. These capacitors present reactance values which are small in comparison to the load and shunt impedances. Therefore the increased capacitive reactance and series resistance with decreased temperature, in the case of the second and third stages, has little or no gain reduction effect. Where the base circuit-blocking capacitors are concerned, the increasing reactance and resistance increases the load impedance on the preceding stages sufficiently to increase power transfer and to compensate for the increased reactance losses. In short, there is a reduction in first-stage gain because of increasing degeneration; a constant or slightly increasing second-stage gain; and a slightly increasing third-stage gain. A contribution to third-stage gain could be made by increasing shunt impedance of the output transformer. At the d-c levels used, a temperature decrease could decrease the collector current sufficiently to cause changes in output transformer parameters. However any changes in the d-c levels of the interstage transformers would be negligible.

PHASE SHIFT AND FREQUENCY RESPONSE

Calculated voltage phase shift is presented in Table III. Methods of calculation may be found in the Appendix. The apparently large phase shifts for the R/C- and direct-coupled amplifiers are due primarily to the effect of the heavily inductive load. Use of a resistive or capacitive load would materially decrease the effective voltage phase shift.

Fig. 5 presents the frequency response curves for the three amplifiers. It should be remembered in comparing the curves that each amplifier is operating at different operating points. The sharp increase in gain of the transformer-coupled amplifier could be attributed to the increasing transformer inductive reactance and decreasing capacitive reactance of the emitter by-pass capacitor. Increasing inductive reactance permits greater power transfer to the respective loads for each stage, while decreasing capacitive reactance decreases degenerative losses. The tapering-off at higher frequencies could conceivably be due to increasing transformer core losses. At 400 cycles, hysteresis and eddy current losses have been considered negligible, but at higher frequencies these losses may acquire importance. Eddy current losses are proportional to the frequency squared while hysteresis losses are approximately directly proportional to frequency. Thus at 20,000 cycles, eddy current losses are approximately 2,500 times greater than at 400 cycles while hysteresis losses are about 50 times greater.

In the R/C- and direct-coupled amplifiers, the gain increase and taparing off effects are less pronounced, since only one transformer is involved. The differences between the points at which the tapering-off trend occurs is attributable to the operating points of the amplifiers. At the operating point of the third stage of the direct-coupled amplier, the inductance of the load transformer is higher

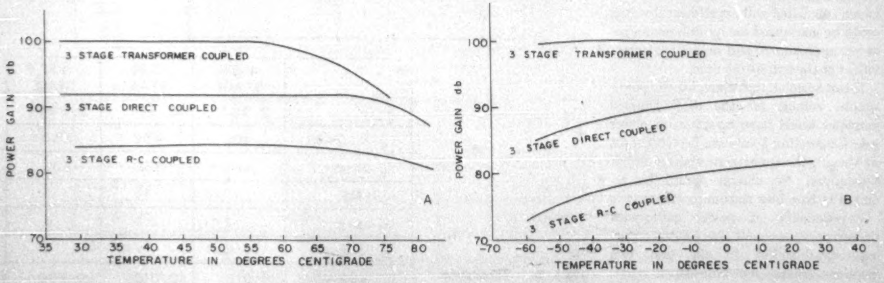


Fig. 4. Variation of power gain with temperature

Table III. Voltage Phase Shift

Type of Amplifier	Calculated Voltage Phase Shift
Transformer-coupled	−157°
R/C-coupled	− 64°
Direct-coupled	− 78°

than at the operating point of the third stage of the R/C-coupled amplifier. Thus, peak gain or maximum transfer condition is approached sooner with the direct coupled amplifier than with the R/C-coupled amplifier. Operation at the same operating points should give quantitatively similar curves.

Number of Components

A quantitive comparison of the components included with each type of amplifier is presented in Table IV. The number of components required for the direct-coupled amplifier is appreciably less than is required by each of the other two amplifiers. No attempt is made to compare production costs of the amplifiers; but, from a strictly general viewpoint, it is apparent from Table IV that cost of components for the direct-coupled amplifier will run considerably less than for the other two types.

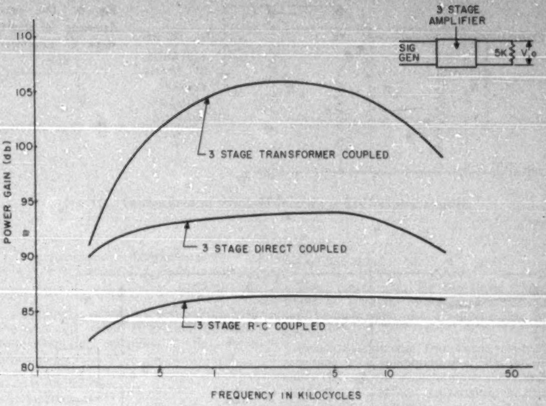


Fig. 5. Frequency response curves

Discussion

In reference to the d-c operating points of the third stages, the possible alternating voltage swing is limited by the d-c collector to emitter voltage. By keeping this voltage at a sufficiently high value, the amplifier is enabled to handle larger input signal levels without clipping. While this is relatively easy to accomplish with R/C and transformer coupling, the problems connected with direct coupling make this possible only by increasing power supply voltage or reducing bias stabilization network resistances. The lower the collector voltage operating point, the lower will be the possible sine wave voltage across the load.

Increase in phase shift with temperature increase should be experienced by all three amplifiers, and will occur almost entirely through the transformers in the circuits. Increased temperatures will change the d-c operating level of the transistors. This change will cause variations in transistor output impedance, transistor input impedance, and transformer primary inductance in such a manner as to increase phase shift through the transformer. This shift should be most pronounced with transformer coupling where three transformers are involved.

The characteristics of the capacitors at increased temperatures are such that little phase shift as a result of capacitance change will be evident. With decreased temperatures, however, capacitance suffers a considerable decrease, producing a phase shift effect that counters, to a certain extent, phase shift changes which may occur through the transformers. Thus, phase shift variations should be less pronounced with decreasing temperature than with increasing temperatures.

Where R/C and direct coupling are concerned the low temperature characteristics of the capacitors adversely affect the low-temperature performance of the amplifiers. Capacitors with improved temperature characteristics would substantially improve the low-temperature performance of R/C- and direct-coupled circuits.

Transformer coupling has the advantages of high-gain and excellent low-temperature performance. Disadvantages include high-temperature performance and phase shift problems which could be alleviated to a certain extent by using transformers rated for higher levels of primary direct current.

Direct coupling has the fewest components and provides good high-temperature performance, little phase shift, and a substantial power gain. Its chief disadvantage in so far as the amplifier treated in this paper is concerned, appears to be the necessity for increasing the power supply voltage in order to maintain the desired bias stabilization network without reducing transistor d-c operating levels below a practical value. However, it is conceivable that a more complete investigation of direct-coupled circuitry could produce a temperature-stable amplifier operating from a 28-volt d-c supply.

The R/C-coupled amplifier could, in most instances, be rated almost on a par with the direct-coupled amplifier. Both R/C and direct coupling display gain deterioration with decreased temperature. This effect is due primarily to the low-temperature characteristics of the tantalytic emitter by-pass capacitors used in the circuits.

General Considerations

This paper has discussed three circuits on the basis of a 10-ohm source impedance, a specified 4-kilohm load impedance and a frequency of 400 cycles. However, different application requirements necessitate additional design considerations.

At frequencies lower than 400 cycles, the value of by-pass and coupling capacitor should be increased to minimize phase shift and degenerative losses. At higher frequencies, it would be primarily from economy and size standpoints that the capacitance values need be decreased. In the transformer-coupled amplifier,

Table IV. Number of Components

Type of Amplifier	Transformers	Capacitors	Resistors	Transistors
Transformer-coupled	3	5	9	3
R/C-coupled	1	5	11	3
Direct-coupled	1	3	7	3

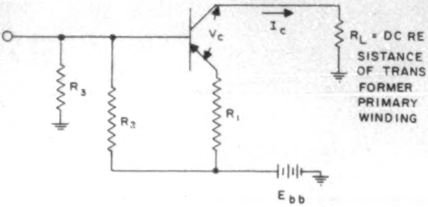


Fig. 6. D-c configuration for single stage of amplification

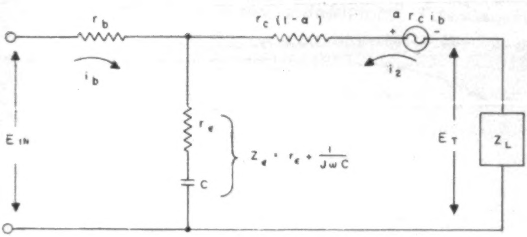


Fig. 7. Equivalent circuit for common emitter amplifier stage

lower or higher frequences, as shown in Fig. 5, could so affect transformer characteristics that different transformers would be required.

For those applications with higher impedances, input impedance in the common emitter configuration may be increased, at the expense of some power gain, by the use of degenerative feedback. For the amplifiers considered in this paper, an unbypassed resistance R placed in the emitter circuit would increase the input impedance by approximately $\alpha/1-\alpha$ times R for $r_L << r_c\,(1-\alpha)$. R would ordinarily be much smaller than the emitter bias resistance, and should therefore cause no noticeable change in the operating point. High values of input impedance with appreciable power gain can be obtained by this, method, especially when a high α-transistor is used. However, the actual input impedance seen by the source would be limited by the values of the base bias resistances.

This limitation could be removed by transformer coupling to the source, since with transformer coupling the base bias resistances do not shunt the transistor input. It may be practical in some instances to use the high input impedance characteristics of a common collector configuration in the first stage with transformer or R/C-coupling to the following common emitter stages.

Load impedances different from the one specified would require recalculation of d-c bias values and a-c performance with the respective d-c and a-c values of the load impedance in the appropriate design and performance equations presented in the Appendix.

Degenerative feedback with the unbypassed emitter circuit resistance could also be used to satisfy lower gain requirements. For example, the second stage of the transformer coupled amplifier of Fig. 1 was calculated to have a power gain of about 32.6 db for the chosen operating point. A power gain of 20 db for this stage would require, from equation 7, an emitter resistance, r^e, of about 1,500 ohms. Since the internal emitter resistance for this stage was held to be about 50 ohms, an external un-by-passed emitter resistance of 1,450 ohms would be needed to obtain the desired gain. The second-stage input impedance would increase, in this case, from the originally calculated 2,390 ohms to about 42 kilohms. This would increase the load impedance for the preceding stage with a resulting increase of the power gain of that stage of less than 1 db. Thus, the over-all power gain of the 3-stage transformer-coupled amplifier is reduced from the original 102.1 db to about 90 db. This principle is also applicable to the R/C- and direct-coupled amplifiers.

Appendix

Design and Performance Calculations

THREE-STAGE TRANSFORMER-COUPLED AMPLIFIER

To assist the engineer newly indoctrinated in transistor applications, the principles and processes associated with the design and performance calculations are demonstrated stage by stage. Bias stabilization resistance values are obtained through application of equations for a single battery supply.[1] The d-c configuration for each stage can be represented by the circuit of Fig. 6. For the first stage

$$R_1 = \frac{\alpha(E_{bc} - V_c - R_L I_c)}{I_c - I_{c0}} \qquad (1)$$

where

$\alpha = 0.967$, arbitrarily selected as representative
$E_{bb} = 28$ volts, d-c power supply
$V_c = 10$ volts direct current
$R_L = 1$ kilohm
$I_c = 0.3$ ma direct current
$I_{c0} = 10\ \mu a$ direct current
$R_1 = 58.7$ kilohms

$$R_2 = \frac{S-1}{\frac{1-S+\alpha S}{R_1} - \frac{I_c - SI_{c0}}{E_{bb}}} \qquad (2)$$

where

$S = 1.3$, arbitrarily selected
$R_2 = 50$ kilohms

$$R_3 = \frac{E_{bb}(S-1)}{I_c - SI_{c0}} \qquad (3)$$

$R_3 = 29.2$ kilohms

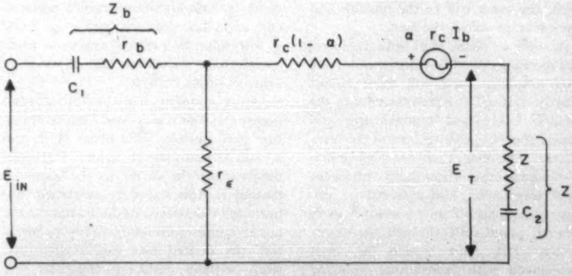


Fig. 8. Equivalent circuit for intermediate stage of transformer-amplifier

Table V. Transformer Parameter Data

Units	Transformer 1	Transformer 2	Transformer 3
R_g kilohms	100	100	17
R_L ohms	2,400−j40	300−j40	50
R_p ohms	1,000	1,000	300
R_s ohms	200	100	10
a^2 ohms	20	20	100
ωL ohms	25,100	20,100	5,000

Since the d-c operating conditions of the first and second stages are the same, the stabilization network resistances will be the same for both stages.

In the third stage, however

$\alpha=0.925$
$V_c=25$ volts
$I_e=5$ ma
$I_{co}=0.3$ ma
$R_L=250$ ohms
$S=1.3$
$E_{bb}=28$ volts

Applying these values to equations 1, 2, and 3

$R_1=345$ ohms
$R_2=122$ ohms
$R_3=1.82$ kilohms

Performance calculations are based on approximate transistor parameter values obtained from average characteristic curves for the operating points concerned. Fig. 1 includes parameter values assumed for the transformer coupled amplifier. Applying the equation for maximum performance[3] to each stage establishes for the third stage a load impedance for maximum gain

$$r_{\max}=r_c(1-\alpha)\sqrt{\frac{r_b(1-\alpha)+r_e}{(r_e+r_b)(1-\alpha)}}=16.5\text{ kilohms} \quad (4)$$

The input impedance corresponding to r_L maximum

$$r_{\text{in}}=\sqrt{(r_e+r_b)\left(r_b+\frac{r_e}{1-\alpha}\right)}=281\text{ ohms} \quad (5)$$

Maximum third stage gain for the specified conditions

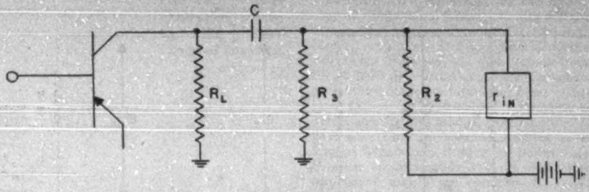


Fig. 10. Configuration of interstage coupling in R/C-coupled amplifier

$$G_{\max}=\frac{\alpha^2 r_c}{[\sqrt{r_e+r_b(1-\alpha)}+\sqrt{(r_e+r_b)(1-\alpha)}]^2}\cong 33\text{ db} \quad (6)$$

Maximum performance figures for both the first and second stages will be same. Application of equations 4 through 6 produces for each of the first two stages

$r_{L\max}=100$ kilohms
$r_{\text{in}}=1{,}620$ ohms
$G_{\max}=39.2$ db

Total maximum power gain of the three stages for peak conditions would be 111.4 db. With the selected load of a 4-kilohm impedance, third-stage gain would be

$$G=\frac{\alpha^2 r_c{}^2 r_L}{[r_c(1-\alpha)+r_L]\times[r_c r_e+r_c r_b(1-\alpha)+r_L(r_e+r_b)]} \quad (7)$$

$G_3\cong 31$ db and

$$r_{\text{in}}=r_b+r_e\left[\frac{r_c+r_L}{r_c(1-\alpha)+r_L}\right] r_{\text{in}_3}\cong 300\text{ ohms} \quad (8)$$

With a 20-to-1 transformer impedance ratio, the third stage r_{in} appears as a r_L of 6 kilohms for the second stage. Applying equations 7 and 8 results in

$G_2\cong 32.6$ db and
$r_{\text{in}_2}=2{,}390$ ohms

The transformer impedance ratio of 20 to 1 makes r_{in_2} appear as a r_L of 43 kilohms to the first stage. Applying equations 7 and 8 results in

$G_1=38.5$ db
$r_{\text{in}_1}=1{,}935$ ohms

Total power gain for the three stages would be 102.1 db representing a voltage gain of 18.8 (10^3) and a current gain of 830 (10^3).

Phase shift in the transformer-coupled amplifier is a combination of the phase shifts through the transformers and phase shifts introduced by capacitance in the emitter and base circuits and the inductive loads of each stage. Fig. 7 presents the equivalent circuit for the first stage.

Analysis by Kirchhoff equations produces the relationship

$$\frac{e_T}{e_{\text{in}}}=\frac{-\alpha r_c Z_L}{r_b(r_c-\alpha r_c+Z_L)+Z_e(Z_L+r_c)} \quad (9)$$

where Z_L = impedance presented by the reactive load, and Z_e consists of emitter resistance and reactance of the emitter circuit by-pass capacitor. For the first stage, Z_L at $400=1{,}000+j25{,}100$ and $Z_e=50-j40$. Applying parameter values from Fig. 1 to equation 9

$$\frac{e_T}{e_{\text{in}}}=-(278\underline{/106^\circ})$$

For the second and third stages, the equivalent circuit becomes that of Fig. 8, where C_1 is the blocking capacitor in the base circuit and C_2 is the emitter bias resistance by-pass capacitor.

Equation 9 is still applicable with $Z_e=V_e$ and Z_L including the capacitive reactance of C_2. Ordinarily this reactance will be so small as to be negligible in comparison to the inductive reactance of the load. The value r_b shall in this instance include the reactance of C_1. The reactance of C_1 in general is small in comparison to other circuit impedances. For the second stage $Z_L=1{,}000+j20{,}100$, $Z_e=50$, and $r_b=1{,}000-j40$. From equation 9

$$\frac{e_T}{e_{\text{in}}}=-(230\underline{/81^\circ})$$

For the third stage, $Z_L\cong 300+j5{,}000$, $Z_e=5$, and $r_b\cong 250-j40$. Applying equation 9

$$\frac{e_T}{e_{\text{in}}}=-(184\underline{/79.1^\circ})$$

to obtain the voltage phase shift through the transformers applying the equation[1]

$$\theta=\tan^{-1}\frac{(R_g+R_p)(R_{\text{in}}+R_s)a^2}{\omega_L[R_g+R_p+(R_{\text{in}}+R_s a^2)]} \quad (10)$$

where

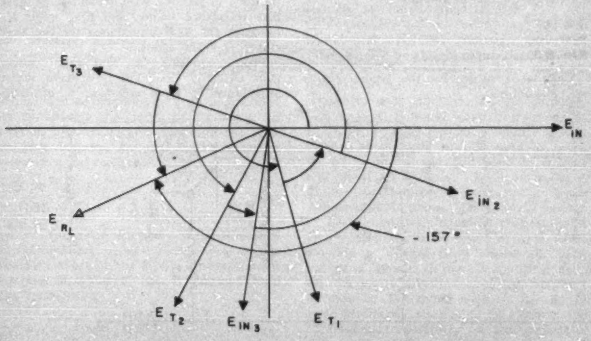


Fig. 9. Voltage phase shift phasors for the transformer-coupled amplifier

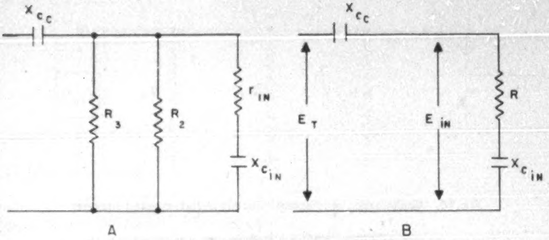


Fig. 11. Simplification of coupling network

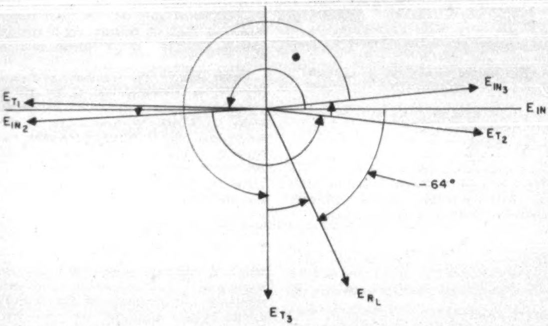


Fig. 12. Voltage phase shift phasors for the R/C-coupled amplifier

R_g = output impedance of preceding stage in ohms
R_{in} = input impedance of following stage in ohms
R_p = primary winding resistance in ohms
R_s = secondary winding resistance in ohms
a = turns ratio
L = primary inductance in henrys
$\omega = 2\pi \times$frequency in cycles per second

Table V presents parameter data for each transformer in the amplifier. R_L includes the effect of the base circuit-blocking capacitors. Solving equation 10 for all three amplifiers

$\theta_{T_1} = 54°$
$\theta_{T_2} = 21°$
$\theta_{T_3} = 42°$

Combining the angles phasorially as in Fig. 9 provides a resultant voltage phase shift for the 3-stage transformer-coupled amplifier of −157 degrees.

THREE-STAGE R/C-COUPLED AMPLIFIER

The same procedure as was used with transformer coupling is used to calculate bias stabilization resistances for the R/C-coupled amplifier. Fig. 7 is again a representation of the d-c configuration for each stage with R_L in this case being the resistor in the collector circuit.

Applying equations 1 through 3 to each stage using related values from Fig. 2 gives, for the first stage, in kilohms:

$R_1 = 11.7$
$R_2 = 6.5$
$R_3 = 8.5$

Calculated values for the second stage are, in kilohms:

$R_1 = 17.5$
$R_2 = 15.5$
$R_3 = 8.5$

and for the third stage, in ohms:

$R_1 = 1{,}750$
$R_2 = 780$
$R_3 = 2{,}300$

Analysis of the performance of the R/C-coupled amplifier involves performance calculations for each stage and correction for stabilization network attenuation. The maximum available gain possible with a 3-stage R/C-coupled amplifier can be found from the relationship

$$G_{max} = \frac{\alpha_1^2\alpha_3^2 r_{c_3}}{4(1-\alpha_1)[r_{e_1}+r_{b_1}(1-\alpha_1)](1-\alpha_3)} \times \left(\frac{\alpha_2}{1-\alpha_2}\right)^2 \qquad (11)$$

where subscripts refer to the stage concerned. Substituting values from Fig. 2 obtains

$G_{max} \cong 87$ db

Load impedance for maximum third stage performance is

$$r_L = r_c(1-\alpha) \qquad (12)$$

$$= 22.5 \text{ kilohms}$$

Third-stage power gain for this load is, from equation 7, 34.8 db. For the selected load of a 4-kilohm impedance, third-stage power gain becomes approximately 31.6 db, with input impedance from equation 8≅307 ohms. This impedance in parallel with the shunt resistances as shown in Fig. 10 becomes the load for the preceding stage. From third-stage bias network values previously calculated and second-stage R_L from Fig. 1, second-stage load impedance becomes

$$r_{L_2} = \frac{1}{\frac{1}{R_L}+\frac{1}{R_3}+\frac{1}{R_2}+\frac{1}{\frac{1}{R_{in_3}}}} \cong 200 \text{ ohms}$$

From equation 7, second-stage power gain $G_2 \cong 22.2$ db, and from equation 8, second-stage input impedance becomes approximately 900 ohms. This input impedance in parallel with second-stage base bias resistance values and first stage R_L becomes the load impedance for the first stage, $r_{L_1} \cong 725$ ohms.

From equation 7, first-stage power gain $G_1 \cong 27.7$ db. The first-stage input impedance is, from equation 8, approxi-

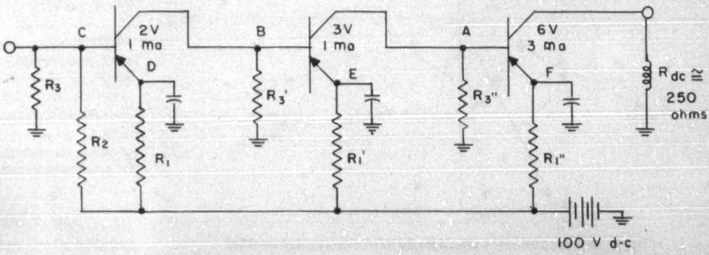


Fig. 13. Direct-coupled circuit used as basis for calculations

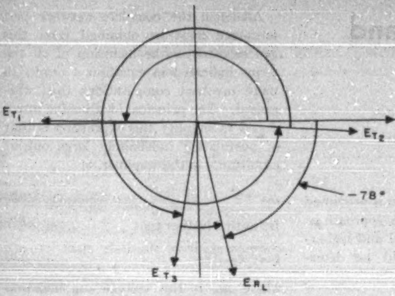


Fig. 14. Voltage phase shift phasors for the direct-coupled amplifier

mately 880 ohms, but the shunting effect of the base bias resistances results in the presentation to the signal source of an impedance, $r_{in_1}\cong 710$ ohms.

Total power gain for the three stages becomes the sum of the corrected power gains for each stage, and for a 22.5-kilohm load, is 84.7 db, while for the selected 4-kilohm load it is 81.5 db.

Phase shift in the R/C-coupled amplifier is a combination of phase shifts introduced by capacitance in the emitter circuits, reactance in the coupling networks, and the reactive loads of each stage. The equation 9 applies equally well to the R/C-coupled amplifier. For the first stage $Z_e=24-j4$ and $Z_L=792-j107$. Z_L includes the reactive effect of emitter circuit by-pass capacitance on input impedance. Using equation 9 and values from Fig. 2, produces for the first stage

$$\frac{\Sigma T_1}{\Sigma_{in_1}}=-(22.4\underline{/-1^\circ})$$

For the second stage $Z_\Sigma=20-j4$ and $ZL=197-j55.5$

$$\frac{\Sigma T_2}{\Sigma_{in_2}}=-(5.6\underline{/-8.2^\circ})$$

For the third stage $Z\Sigma=5-jr$ and $Z_L=250+j6{,}260$

$$\frac{\Sigma T_3}{\Sigma_{in_3}}=-(246\ \underline{/85^\circ})$$

Phase shift in the transformer coupling to the load is from equation 10 equal to 26 degrees, obtained using a primary inductance value of 2.5 henrys related to the 4-ma third-stage d-c operating level. Voltage phase shift relationship through the capacitive coupling network is obtained by simplifying the circuit of Fig. 11(A) to that of Fig. 11(B) and obtaining the relationship between Σ_T and Σ_{in} in terms of resistance and reactance. Thus,

$$\frac{\Sigma_{in}}{\Sigma_T}=\frac{R-jX_{c_{in}}}{R-j(X_{c_{in}}+X_{c_c})} \qquad (13)$$

where X_{c_c} is the reactance of the 10-microfarad coupling capacitor and $X_{c_{in}}$ is the reactive effect of the 20 *μfd* emitter circuit by-pass capacitance. Z_{in} is obtained from the relationship

$$Z_{in}=r_b+\frac{1}{1-\alpha}\left(r_e+\frac{1}{j\omega c_{in}}\right) \qquad (14)$$

and

$$R-jX_{c_{in}}=\frac{1}{\frac{1}{Z_{in}}+\frac{1}{R_2}+\frac{1}{R_3}} \qquad (15)$$

Applying equations 14, 15, and 13 to the coupling between the first and second stages produces the relationship

$$\frac{\Sigma_{in_2}}{\Sigma_{T_1}}=0.992\ \underline{/2.9^\circ}$$

and repeating the procedure for the coupling between the second and third stages produces

$$\frac{\Sigma_{in_3}}{\Sigma_{T_2}}=0.963\underline{/11^\circ}$$

Combining all angles phasorially as in Fig. 12 provides a resultant voltage phase shift for the 3-stage R/C-coupled amplifier of −64 degrees.

THREE-STAGE DIRECT-COUPLED AMPLIFIER

Bias resistances are approximated on the basis of desired transistor d-c operating points and power supply voltage. Fig. 13 shows that calculations are started from the third stage. With 3 ma in the collector circuit, a 0.75-volt drop will occur across the 250-ohm d-c resistance of the transformer primary. With a 6-volt drop across the transistor, 93.25 volts would remain to be dropped across R_1''. At approximately a 3-ma emitter current, this would make $R_1''\cong 31$ kilohms. With the voltage at *A* almost equivalent to the voltage at *F*, essentially 7 volts must be dropped across R_2''. With a 1-ma second-stage collector current, $R_2''\cong 7$ kilohms. Ten volts at *E* would leave 90 volts across R_1'. At 1 ma, $R_1=90$ kilohms. With voltage at *B* essentially the same as at *E*, R_2 for 1-ma first-stage collector current will equal 10 kilohms. With 12 volts at *D*, R_1 must handle 88 volts and at 1 ma will equal 88 kilohms. To maintain voltage at *C* approximately equal to that at point *D*, ratio of R_2 to R_3 must be such as to provide a 12-volt drop across R_3. The values of R_2 and R_3 have a shunting effect on the input impedance of the amplifier as seen by the signal source. R_2 and R_3 were arbitrarily selected as 88 kilohms and 12 kilohms respectively for the purpose of this application.

Performance calculations for the direct-coupled amplifier follow, to a certain extent, the pattern of those for the R/C coupled amplifier. The maximum available gain, from equation 11, with values from Fig. 3 $\cong$89 db. Load impedance for maximum third-stage performance, from equation 12, is 18.75 kilohms. Third-stage power gain for this load is, from equation 7, 34.8 db. For the selected load of a 4-kilohm impedance, third-stage power gain becomes approximately 31.7 db with input impedance and, from equation 8, $\cong$250 ohms. This impedance in parallel with the base bias resistance becomes the load for the preceding stage. With the third-stage bias resistance previously established at 7 kilohms, the load impedance for the second stage should be 240 ohms. The power gain of the second stage, from equation 7, $G_2\cong$ 24.1 db, and second-stage input impedance, becomes approximately 780 ohms. This impedance is parallel with the second-stage base bias resistance of 10 kilohms, and provides a load impedance for the first stage of about 730 ohms. First-stage power gain for this load, $G_1\cong 29.4$ db. Input impedance is about 730 ohms. Including the shunting effect of the first-stage base bias resistances would result in an impedance of about 700 ohms being presented to the signal source. Total power gain for the three stages becomes the sum of the power gains of each stage and, for a 18.75-kilohm load, is 88.3 db, while for the selected 4-kilohm load it is about 85.2 db.

Phase shift in the direct-coupled amplifier is a combination of phase shifts introduced by the capacitance in the emitter circuits and the reactive loads of each stage. Equation 9 and Fig. 8 are used to obtain the voltage phase shift of each stage. For the first stage, $Z_e=15-j4$ and $Z_L=718-j104$. With values from Fig. 3

$$\frac{\Sigma_{T_1}}{\Sigma_{in_2}}=-(24.5\ \underline{/-1^\circ})$$

For the second stage

$$\frac{\Sigma_{T_2}}{\Sigma_{in_3}}=-(9.55\ \underline{/-2.5^\circ})$$

For the third stage, $Z_e=5-j4$. At the third-stage d-c operating level of 3 ma the transformer primary inductance is 3 henrys and primary impedance Z_L is about $250+j7{,}530$.

$$\frac{\Sigma_{T_3}}{\Sigma_{in_3}}=-(340\ \underline{/82^\circ})$$

Phase shift in the transformer coupling to the load, from equation 10 is approximately 23 degrees for the primary inductance of 3 henrys. Combining all angles phasorially as in Fig. 14 provides a resultant voltage phase shift for the 3-stage direct-coupled amplifier of −78 degrees.

References

1. PRINCIPLES OF TRANSISTOR CIRCUITS (book), R. F. Shea. John Wiley & Sons, Inc., New York, N. Y., 1953, chaps. 5, 6.

2. REFERENCE DATA FOR RADIO ENGINEERS (book), Federal Telephone and Radio Corporation. Knickerbocker Printing Corporation, New York, N.Y., 1949, third edition, chap. 11.

Analogue Computer Synthesis and Error Matrices

P. M. HONNELL
MEMBER AIEE

R. E. HORN
ASSOCIATE MEMBER AIEE

Synopsis: An evaluation of the actual solutions exhibited by an electronic analogue computer, and thus a measure of its accuracy and stability, requires the analysis of a physical system represented by a linear differential equation system of higher order than that of the mathematical problem being solved. To this situation the powerful and systematic methods of matrix analysis are applied, yielding expressions from which the degree of accuracy which may be attained in computer solutions may be determined and, in particular, evaluating the all-important problems of stability of the electronic networks.

AN ELECTRONIC analogue computer programmed for a system of linear differential equations ideally exhibits solutions of the mathematical problem. Such a performance necessarily can be only approximated since idealizations cannot actually be achieved physically, for the networks in an electronic analogue machine are far more complex than the mathematical problem equations alone imply. To analyze so involved an electronic network systematically, a sufficiently general method of analysis is required.

The object of this paper is to apply the powerful and systematic methods of matrix analysis to determine more precisely the actual performance of the analogue machine. Clearly, analysis considering all physical parameters in the computer, while feasible and straightforward, is unnecessarily involved. This paper is therefore constrained to approximations which may be considered as first-order, and which yield a measure of the magnitude and nature of the approximations of the true mathematical solutions of the given problem, and to the all-important problem of stability. Furthermore, linearity of the system is presumed in this analysis. The matrix approach is, however, completely general and higher-order approximations could be determined if desired.

Basic Electronic Computer Matrices

The basic electronic computer networks are shown in Fig. 1(A) and (B), in which μ_1 and μ_2 represent multistage electronic amplifiers and the y_{ii} are primitive dinodes, i.e., capacitors and conductances or their combinations. These basic computer configurations are analyzed and synthesized on a common-node coordinate basis.

The n-stage amplifiers μ_1 and μ_2 are assumed to have admittance matrices $[y]$ of the following form

$$[y]=\begin{bmatrix} 0 & 0 \\ (-1)^{n+1}\dfrac{g_{m1}g_{m2}\cdots g_{m,n-1}g_{mn}}{(g_{p1}+y_{p1})(g_{p2}+y_{p2})\cdots(g_{p,n-1}+y_{p,n-1})} & (g_{pn}+y_{pn}) \end{bmatrix} \tag{1}$$

in which the g_m:*s* and g_p:s are the mutual and plate conductances respectively, while the y_P:*s* represent the frequency-dependent dinode interstage networks. This equation may be abbreviated as

$$[y]=\begin{bmatrix} 0 & 0 \\ \mu_0(d)g_{mn} & (g_{pn}+y_{pn}) \end{bmatrix} \tag{2}$$

in which the frequency dependence and voltage amplification of the interstage networks is represented by $\mu_0(d)$, where $d=d/dt$. In this analysis, n is considered an odd number.

The y-matrix of the entire network of Fig. 1(A), or 1(B), may be obtained from the sum of equation 2 and the y-matrices for the dinode admittances y_{rk} which have the form

$$[y]=\begin{bmatrix} y_{rk} & -y_{rk} \\ -y_{rk} & y_{rk} \end{bmatrix} \tag{3}$$

Conventional matrix operations will yield the matrices of the complete networks. For Fig. 1(A) this may be readily verified as

$$[y]=\begin{bmatrix} \dfrac{y_{11}(\mu_0(d)\times g_{mn}+(g_{pn}+y_{pn})+y_{21})}{(g_{pn}+y_{pn})+y_{11}+y_{21}} & \dfrac{-y_{11}y_{21}}{(g_{pn}+y_{pn})+y_{11}+y_{21}} \\ \dfrac{y_{21}(\mu_0(d)\times g_{mn}-y_{11})}{(g_{pn}+y_{pn})+y_{11}+y_{21}} & \dfrac{y_{21}((g_{pn}+y_{pn})+y_{11})}{(g_{pn}+y_{pn})+y_{11}+y_{21}} \end{bmatrix} \tag{4}$$

Although the complete network performance may be obtained from this general matrix, the inclusion of all the terms indicated in equation 4 results in more involved computations than warranted. The principal idealization introduced to simplify this expression is that of postulating indefinitely large output admittance in the amplifier, or

$$g_{pn}\rightarrow\infty \tag{5}$$

By noting further that

$$g_{mn}=g_{pn}\times\mu_{0n} \tag{6}$$

where μ_{0n} is the voltage amplification factor of the last stage, equation 4 may be approximated by

$$[y]=\begin{bmatrix} y_{11}(\mu_0(d)\times\mu_{0n}+1) & 0 \\ y_{21}\times\mu_0(d)\times\mu_{0n} & y_{22} \end{bmatrix} \tag{7}$$

By further defining for compactness

$$\mu_0(d)\times\mu_{0n}\equiv\mu_1(d)\equiv\mu_1 \tag{8}$$

equation 7 becomes

$$[y]=\begin{bmatrix} y_{11}\times(\mu_1+1) & 0 \\ y_{21}\times\mu_1 & y_{22} \end{bmatrix} \tag{9}$$

It is convenient to rewrite equation 9 in the following manner

$$[y]=\begin{bmatrix} y_{11}\times\mu_1 & 0 \\ y_{21}\times\mu_1 & 0 \end{bmatrix}+\begin{bmatrix} y_{11} & 0 \\ 0 & y_{22} \end{bmatrix} \tag{10}$$

The reason for the partitioning of equation 9 into the form of equation 10 (correspondingly for equation 11) is that as will be shown subsequently, a transformation of the first part will place it in 1-to-1 correspondence with the mathematical problem, leaving the second part as the principal source of error in the computer system.

The network of Fig. 1(B) may be analyzed similarly, and its y-matrix verified as

$$[y]=\begin{bmatrix} 0 & y_{12}\times\mu_2 \\ 0 & y_{22}\times\mu_2 \end{bmatrix}+\begin{bmatrix} y_{12} & 0 \\ 0 & y_{22} \end{bmatrix} \tag{11}$$

Although it is possible to formulate an electronic network synthesis of mathematical problems based upon these simple networks, it is advantageous to consider their generalization, Fig. 2.

Paper 56-49, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted August 1, 1955; made available for printing November 17, 1955.

P. M. HONNELL is with Washington University, St. Louis, Mo., and R. E. HORN is with the Westinghouse Electric Corporation, Baltimore, Md.

This paper summarizes some of the researches reported in a doctoral dissertation submitted by the junior author to the Sever Institute of Technology of Washington University.

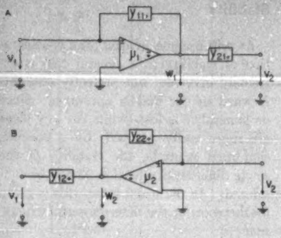


Fig. 1. Basic electronic synthesis networks

The representation in Fig. 2 includes amplifiers with bipolar output polarities, and additional dinode admittances y_{sk}. The admittance matrix of this generalized electronic network, following the pattern of equations 10 and 11, then reads

$$[y]=\begin{bmatrix} 0 \,.\, 0 & (y_{1k+}-y_{1k-})\mu_k & 0 \,.\, 0 \\ 0 \,.\, 0 & (y_{2k+}-y_{2k-})\mu_k & 0 \,.\, 0 \\ \cdots & \cdot \quad \cdot \quad \cdot & \cdots \\ 0 \,.\, 0 & (y_{kk+}-y_{kk-})\mu_k & 0 \,.\, 0 \\ \cdots & \cdot \quad \cdot \quad \cdot & \cdots \\ 0 \,.\, 0 & (y_{rk+}-y_{rk-})\mu_k & 0 \,.\, 0 \\ 0 \,.\, 0 & (y_{sk+}-y_{sk-})\mu_k & 0 \,.\, 0 \end{bmatrix}+\begin{bmatrix} (y_{1k+}+y_{1k-}) & . & 0 & . & 0 \\ 0 & . & 0 & . & 0 \\ . & . & . & . & . \\ 0 & . & (y_{kk+}+y_{kk-}) & . & 0 \\ . & . & . & . & . \\ 0 & . & 0 & . & 0 \\ 0 & . & 0 & . & (y_{sk+}+y_{sk-}) \end{bmatrix} \tag{12}$$

The first part of equation 12 is called the ideal synthesis matrix, because it can be placed in 1-to-1 correspondence with the mathematical problem to be solved, after a transformation. The second part is called the network error matrix; this is of diagonal form and one of the sources of error in the computer.

In addition to the network error matrix, additional matrix terms indicating the errors due to approximations of ideal current sources by equivalent voltage sources and series impedances, and due to parasitic parameters in the physical network may also be important. The terms in the equivalent source error matrix appear on the principal diagonal, and are primarily conductances. Parasitic elements may be present in all the matrix entries, but those which predominate are usually capacitive terms also on the principal diagonal, and often due to shielded input leads to the amplifiers.

Thus, the matrix admittance equations for a general ($s \times s$) order computing network may be represented by the following:

$$([1_y]+[2_y]+[3_y]+[4_y])\begin{bmatrix} v_1 \\ v_2 \\ . \\ v_k \\ . \\ v_r \\ v_s \end{bmatrix}=\begin{bmatrix} j_1' \\ j_2' \\ . \\ j_k' \\ . \\ j_r' \\ j_s' \end{bmatrix} \tag{13}$$

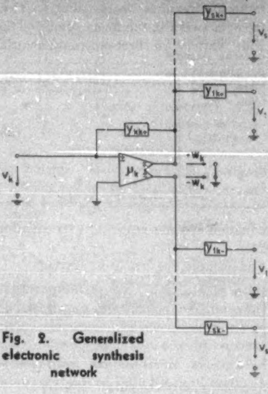


Fig. 2. Generalized electronic synthesis network

in which the ideal synthesis matrix now is

$$[1_y]=\begin{bmatrix} (y_{11+}-y_{11-})\mu_1 & . \; . & (y_{1k+}-y_{1k-})\mu_k & . \; . & (y_{1s+}-y_{1s-})\mu_s \\ (y_{21+}-y_{21-})\mu_1 & . \; . & (y_{2k+}-y_{2k-})\mu_k & . \; . & (y_{2s+}-y_{2s-})\mu_s \\ . & . \; . & . & . \; . & . \\ (y_{k1+}-y_{k1-})\mu_1 & . \; . & (y_{kk+}-y_{kk-})\mu_k & . \; . & (y_{ks+}-y_{ks-})\mu_s \\ . & . \; . & . & . \; . & . \\ (y_{r1+}-y_{r1-})\mu_1 & . \; . & (y_{rk+}-y_{rk-})\mu_k & . \; . & (y_{rk+}-y_{rs-})\mu_s \\ (y_{s1+}-y_{s1-})\mu_1 & . \; . & (y_{sk+}-y_{sk-})\mu_k & . \; . & (y_{ss+}-y_{ss-})\mu_s \end{bmatrix} \tag{14}$$

and the network error matrix is

$$[2_y]=\begin{bmatrix} \sum_{t=1}^{s}(y_{1t+}+y_{1t-}) & . & 0 & . & 0 \\ 0 & . & 0 & . & 0 \\ . & . & . & . & . \\ 0 & . & \sum_{t=1}^{s}(y_{kt+}+y_{kt-}) & . & 0 \\ . & . & . & . & . \\ 0 & . & 0 & . & 0 \\ 0 & . & 0 & . & \sum_{t=1}^{s}(y_{st+}+y_{st-}) \end{bmatrix} \tag{15}$$

The equivalent source error matrix $[3_y]$ and the predominant parasitic (capacitive) error matrix $[4_y]$ are both diagonal in form, as previously stated, and may or may not be significant, depending upon the particular computing machine. Finally, the right-hand column of j_s':s in equation 13 represents the current-source equivalents constructed from intrinsic voltage-sources and series impedances.

Applying to equation 13 the transformation

$$[v]=C[-w] \tag{16}$$

where

$$C=\begin{bmatrix} \frac{1}{\mu_1} & 0 & . \; 0 & . \; 0 & 0 \\ 0 & \frac{1}{\mu_2} & . \; 0 & . \; 0 & 0 \\ . & . & . \; . & . \; . & . \\ 0 & 0 & . \; \frac{1}{\mu_k} & . \; 0 & 0 \\ . & . & . \; . & . \; . & . \\ 0 & 0 & . \; 0 & . \; \frac{1}{\mu_r} & 0 \\ 0 & 0 & . \; 0 & . \; 0 & \frac{1}{\mu_s} \end{bmatrix} \tag{17}$$

results in

$$([0_y]+[2_y]C+[3_y]C+[4_y]C)\begin{bmatrix} -w_1 \\ -w_2 \\ . \\ -w_k \\ . \\ -w_r \\ -w_s \end{bmatrix}=\begin{bmatrix} j_1' \\ j_2' \\ . \\ j_k' \\ . \\ j_r' \\ j_s' \end{bmatrix} \tag{18}$$

It should be noted that the *-w*:*s* represent the actual amplifier output voltages.

In equation 18 the term

$$[0_y]=[1_y]C \tag{19}$$

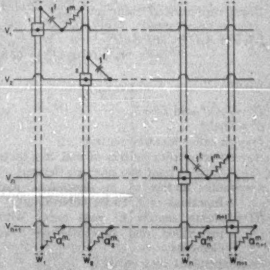


Fig. 3 (right). Ideograph of synthesis network according to equation 25, exhibiting solution to equation 23 which will be shown to correspond exactly to the mathematical problem matrix, now reads

$$[{}^0_y]=\begin{bmatrix} (y_{11+}-y_{11-}) & \cdot\ \cdot & (y_{1k+}-y_{1k-}) & \cdot\ \cdot & (y_{1s+}-y_{1s-}) \\ (y_{21+}-y_{21-}) & \cdot\ \cdot & (y_{2k+}-y_{2k-}) & \cdot\ \cdot & (y_{2s+}-y_{2s-}) \\ \cdot & \cdot\ \cdot & \cdot & \cdot\ \cdot & \cdot \\ (y_{k1+}-y_{k1-}) & \cdot\ \cdot & (y_{kk+}-y_{kk-}) & \cdot\ \cdot & (y_{ks+}-y_{ks-}) \\ \cdot & \cdot\ \cdot & \cdot & \cdot\ \cdot & \cdot \\ (y_{r1+}-y_{r1-}) & \cdot\ \cdot & (y_{rk+}-y_{rk-}) & \cdot\ \cdot & (y_{rs+}-y_{rs-}) \\ (y_{s1+}-y_{s1-}) & \cdot\ \cdot & (y_{sk+}-y_{sk-}) & \cdot\ \cdot & (y_{ss+}-y_{ss-}) \end{bmatrix} \tag{20}$$

and no longer exhibits the amplifier gains the μ:s.

The numerically small entries in the matrices $[{}^2_y]C$, $[{}^3_y]C$, $[{}^4_y]C$, resulting from large magnitude μ_k, do not assure stability of the computer network. In a qualitative sense, the fact that these terms are "small" usually results in "high-frequency" stability problems. Clearly, the mathematical problem stability, indicated by the properties of $[{}^0_y]$ alone, is entirely distinct from the electronic network stability, which is determined by all four matrix terms in equation 18.

Furthermore, each allowable transformation of the mathematical problem and synthesis matrices represents a new and distinct electronic network due to the error matrices, although exhibiting, ideally, the same mathematical solution. Since the stability of the electronic network depends on all terms of equation 18 and not on the synthesis matrix $[{}^0_y]$ alone, each such transformation may yield either a stable or an unstable electronic network. This acts as an undesired constraint on the type of equivalent mathematical system which may be synthesized on any given computer.

Mathematical Problem Matrices

It has been shown[1] that the synthesis matrix $[{}^0_y]$, equation 20, may be placed in 1-to-1 correspondence with the matrix representation of a differential equation system. This means, briefly, that programming an electronic analogue computer for a system of differential equations, represented compactly as

$$(A_1D^0+A_2D+\ldots+A_nD^{n-1}+A_{n+1}D^n)x=f(t) \tag{21}$$

where

$D^n=I\times d^n$ and $D^0=I$

$d^n=d^n/dt^n$

$I=m$th order identity matrix

$A_i=$an mth order matrix of real constants with at least one nonzero element

$x=$column matrix of m dependent real functions of t, to be determined

$f(t)=$column matrix of mth order with prescribed real functions of t as elements

$t=$independent real variable

in reality requires synthesizing an electronic network in the image of one of the wide variety of first-order differential equation systems equivalent to equation 21.

For example, the first-order system

$$\begin{bmatrix} D & -I & 0 & \cdot\ \cdot & 0 & 0 \\ 0 & D & -I & \cdot\ \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot\ \cdot & D & -I \\ A_1 & A_2 & A_3 & \cdot\ \cdot & A_n & A_{n+1} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ 0 \\ f(t) \end{bmatrix} \tag{22}$$

in which $x_i=D^{i-1}x$, $i=1,\ldots,n,n+1$ yields the solutions to the given problem equation 21 which represents a system in m variables, with at least one of the equations of the nth order.

A single nth order equation such as

$$(a_1+a_2d+\ldots+a_nd^{n-1}+a_{n+1}d^n)x=0 \tag{23}$$

may similarly be represented as

$$\begin{bmatrix} d & -1 & 0 & 0 & \cdot & 0 & 0 \\ 0 & d & -1 & 0 & \cdot & 0 & 0 \\ 0 & 0 & d & -1 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & d & -1 \\ a_1 & a_2 & a_3 & a_4 & \cdot & a_n & a_{n+1} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ x_n \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ 0 \\ 0 \end{bmatrix} \tag{24}$$

The synthesis matrix $[{}^0_y]$ for the problem equation 23 in this instance reads

$$\begin{bmatrix} c_{11}d & -g_{12} & 0 & \cdot\ \cdot & 0 & 0 \\ 0 & c_{22}d & -g_{23} & \cdot\ \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot\ \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot\ \cdot & c_{nn}d & -g_{n,n+1} \\ g_{n+1,1} & g_{n+1,2} & g_{n+1,3} & \cdot\ \cdot & g_{n+1,n} & g_{n+1,n+1} \end{bmatrix} \begin{bmatrix} -w_1 \\ -w_2 \\ \cdot \\ -w_n \\ -w_{n+1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ 0 \\ 0 \end{bmatrix} \tag{25}$$

in which the c:s are capacitances, and the g:s conductances located in the network as indicated by their ordering numerical subscripts. Initially, the actual numerical magnitude of these dinode elements (usually in the microfarad and micromho range) are determined by the numerical coefficients of the original differential equation and the transformations indicated. However, as shown in the reference,[1] there is a wide latitude for the magnitude of these coefficients, the choice depending actually upon the characteristics of the particular computer.

Although the electronic network synthesis is inherent in the matrix notation, a symbolic network representation of equation 25 is denoted in the ideograph Fig. 3, as a practical convenience.

Stability

As will become evident, the consideration of stability in a general case is extremely involved, but perfectly straightforward by the matrix approach. Since no generality is lost (while the key ideas are more easily grasped) by a specific, simplified example, the stability of the single differential equation 23 will be considered. In the usual sense, this is given by the roots of the determinantal expression

$$a_1+a_2\lambda+\ldots+a_n\lambda^{n-1}+a_{n+1}\lambda^n=0 \tag{26}$$

which may be abbreviated to read

$$\Delta(\lambda)=0 \tag{27}$$

A single λ-root of equation 26 with a positive real part indicates that the mathematical solution, in which $x_i=X_ie^{\lambda t}$ represents a typical term, is unstable; conversely, if all roots of equation 26 have negative real parts the mathematical problem solution is stable. These considerations, however, do not guarantee the stability of the electronic computer network synthesized to exhibit the solution to the problem.

To simplify the discussion of the stability of the synthesized electronic network, it is convenient to consider the dinode network parameters shown in Fig. 3 to be in farads and mhos. Thus, equation 25 will read

$$\begin{bmatrix} d & -1 & \cdot\ \cdot & 0 & 0 \\ 0 & d & \cdot\ \cdot & 0 & 0 \\ \cdot & \cdot & \cdot\ \cdot & \cdot & \cdot \\ 0 & 0 & \cdot\ \cdot & d & -1 \\ a_1 & a_2 & \cdot\ \cdot & a_n & a_{n+1} \end{bmatrix} \begin{bmatrix} -w_1 \\ -w_2 \\ \cdot \\ -w_n \\ -w_{n+1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ 0 \\ 0 \end{bmatrix} \tag{28}$$

and the mathematical and synthesis matrices become identical numerically.

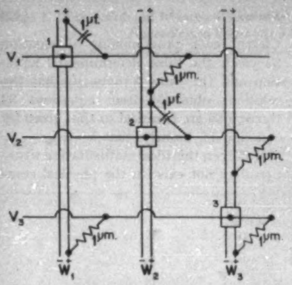


Fig. 4. Ideograph of synthesis network exhibiting solution of $(1+d^2)x=0$ on GEDA, according to equation 34

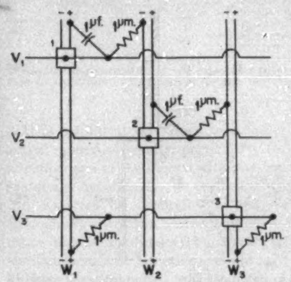


Fig. 5. Ideograph of synthesis network exhibiting solution of $(1+d^2)x=0$ on experimental computer, according to equation 35

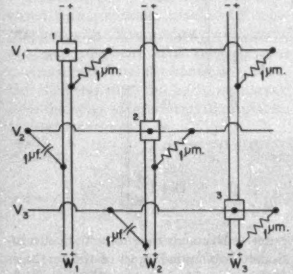


Fig. 6. Ideograph of synthesis network exhibiting solution of $(1+d^2)x=0$ on experimental computer, according to equation 36

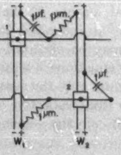


Fig. 7. Ideograph of synthesis network exhibiting solution of $(1+d^2)x=0$ on experimental computer, according to equation 37

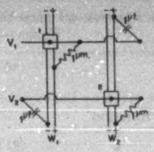


Fig. 8 (left). Ideograph of synthesis network exhibiting solution of $(1+d^2)x=0$ on experimental computer, according to equation 38

Fig. 9 (right). Ideograph of synthesis network exhibiting solution of $(1+d^2)x=0$ on experimental computer, according to equation 39

GEDA	(n) self-d
(n+1) self-d	(n) mutual-d
(n+1) mutual-d	(n+1 mixed-d

Corresponding to the ideal synthesis matrix shown in equation 28, the complete electronic network admittance matrix comprises four distinct matrix terms, since the electronic network is more complicated than the mathematical matrix would indicate. For simplicity and without loss of generality, only the first two terms of equation 18 will be considered in the determination of the nature of the electronic network stability (the other terms can easily be included in a numerical case, but are rather clumsy if included in the following general expressions). The two terms are the synthesis matrix shown above, or

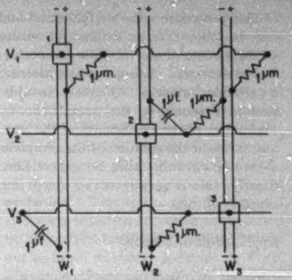
$$[0_y]=\begin{bmatrix} d & -1 & \cdot & 0 & 0 \\ 0 & d & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & d & -1 \\ a_1 & a_2 & \cdot & a_n & a_{n+1} \end{bmatrix} \tag{29}$$

and the network error matrix

$$[2_y]=\begin{bmatrix} (d+1)\frac{1}{\mu_1} & 0 & \cdot & 0 & 0 \\ 0 & (d+1)\frac{1}{\mu_2} & \cdot & 0 & 0 \\ \cdot & & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & (d+1)\frac{1}{\mu_n} & 0 \\ 0 & 0 & \cdot & 0 & \sum_{i=1}^{n+1}|a_i|\,\frac{1}{\mu_{n+1}} \end{bmatrix} \tag{30}$$

The stability of the electronic network, under these simplifications, is determined by the roots of the determinant of the sum of the matrices 29 and 30, that is by

$$\left\|\begin{bmatrix} \lambda+(\lambda+1)\frac{1}{\mu_1} & -1 & \cdot & 0 & 0 \\ 0 & \lambda+(\lambda+1)\frac{1}{\mu_2} & \cdot & 0 & 0 \\ \cdot & & \cdot & & \cdot \\ 0 & 0 & \cdot & \lambda+(\lambda+1)\frac{1}{\mu_n} & -1 \\ a_1 & a_2 & \cdot & a_n & a_{n+1}+\sum_{i=1}^{n+1}|a_i|\frac{1}{\mu_{n+1}} \end{bmatrix}\right\|=0 \tag{31}$$



The λ-roots of equation 31 are now the exponents of the exponential terms of the solution of the differential equation of the complete electronic network. Reduction of the determinant 31 yields the compact expression

$$\Delta\left(\lambda+(\lambda+1)\frac{1}{\mu}\right)+\sum_{i=1}^{n+1}|a_i|\frac{1}{\mu}\left(\lambda+(\lambda+1)\frac{1}{\mu}\right)^n=0 \tag{32}$$

in which it should be noted that $\Delta()$ is defined by equation 27, and all μ_i are taken as equal for simplicity.

It is evident that equation 32 is far more complex than the determinantal equation 27 of the original differential equation 23. Furthermore, the nature of the λ-roots of equation 32, which determine the stability of the complete electronic network depends not only upon the original problem but also upon the choice of first-order system used to represent it, and finally upon the amplifier characteristics, the $\mu_i(d)$.

This procedure is perfectly general and may be followed to investigate the nature of the λ-roots of any of the various electronic networks chosen for synthesizing a particular problem. Of course, in addition to determining the actual λ-roots, evaluation of the coefficients of the various terms of the solution of the complete electronic network could be undertaken. Clearly, this is very involved except for simplest of cases.

Application of Theory

Application of the foregoing theory can be illustrated by the simple differential equation

$$(1+d^2)x=0 \tag{33}$$

Of the indefinitely large variety of representations in which this equation may be programmed for synthesis on an analogue computer, the following six equivalent first-order systems may be classed as typical. These examples will also illustrate the method of application of the synthesis and error matrices to the study of the stability of electronic computer networks. It is important to note that most of the electronic networks corresponding to the synthesis matrices of the following first-order systems are not accessible by the techniques heretofore available in the art.

The first-order systems chosen are given in equations 34 through 39. The corresponding computer programming of these examples is illustrated by the ideographs in Figs. 4 through 9 respectively. The descriptive titles of the various synthesis matrices refer to the location of the differential terms, the *d:s*, represented in the ideographs by the capacitance symbols.

GEDA L3

$$\textit{GEDA}\quad \begin{bmatrix} d & 1 & 0 \\ 0 & d & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -w_1 \\ +w_2 \\ -w_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{34}$$

EXPERIMENTAL COMPUTER

$$(n+1)\ \text{self-}d\quad \begin{bmatrix} d & -1 & 0 \\ 0 & d & -1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -w_1 \\ -w_2 \\ -w_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{35}$$

$$(n+1)\ \text{mutual-}d\quad \begin{bmatrix} 1 & 0 & 1 \\ -d & 1 & 0 \\ 0 & -d & 1 \end{bmatrix} \begin{bmatrix} -w_1 \\ -w_2 \\ -w_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{36}$$

$$(n)\ \text{self-}d\quad \begin{bmatrix} d & -1 \\ 1 & d \end{bmatrix} \begin{bmatrix} -w_1 \\ -w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{37}$$

$$(n)\ \text{mutual-}d\quad \begin{bmatrix} 1 & d \\ -d & 1 \end{bmatrix} \begin{bmatrix} -w_1 \\ -w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{38}$$

$$(n+1)\ \text{mixed-}d\quad \begin{bmatrix} 1 & 0 & 1 \\ 0 & d & -1 \\ -d & 1 & 0 \end{bmatrix} \begin{bmatrix} -w_1 \\ -w_2 \\ -w_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{39}$$

The solution to equation 33 may be written in the form

$$x=X_1\epsilon^{\lambda_1 t}+X_2\epsilon^{\lambda_2 t} \tag{40}$$

in which the *λ:s* are given by the roots of the determinantal equation

$$|1+\lambda^2|=0 \tag{41}$$

which yields for the λ-roots

$$\lambda=\pm i \tag{42}$$

Clearly, each programming of equation 33 yields the same determinantal equation and the same λ-roots, so that the solutions obtained from equations 34 through 39 are identical to that given by equation 41, namely, that $\lambda=\pm i$.

However, this ideal mathematical situation does not exist in the physical computer. Although the mathematical solution 40 is bounded, most of the synthesis matrices yield unstable electronic networks with conventional computers. To reiterate, the stability of the computer network is not indicated by the synthesis matrices alone.

$$\textit{GEDA}\quad \left\| \overset{\text{Synthesis}}{\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 1 & 0 & 1 \end{bmatrix}} + \overset{\text{Error}}{\begin{bmatrix} (\lambda+1)\frac{1}{\mu_a} & 0 & 0 \\ 0 & (\lambda+1)\frac{1}{\mu_b} & 0 \\ 0 & 0 & (2)\frac{1}{\mu_c} \end{bmatrix}} \right\| = 0 \tag{43}$$

$$(n+1)\ \text{self-}d\quad \left\| \begin{bmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ 1 & 0 & 1 \end{bmatrix} + \begin{bmatrix} (\lambda+1)\frac{1}{\mu} & 0 & 0 \\ 0 & (\lambda+1)\frac{1}{\mu} & 0 \\ 0 & 0 & (2)\frac{1}{\mu} \end{bmatrix} \right\| = 0 \tag{44}$$

$$(n+1)\ \text{mutual-}d\quad \left\| \begin{bmatrix} 1 & 0 & 1 \\ -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \end{bmatrix} + \begin{bmatrix} (2)\frac{1}{\mu} & 0 & 0 \\ 0 & (\lambda+1)\frac{1}{\mu} & 0 \\ 0 & 0 & (\lambda+1)\frac{1}{\mu} \end{bmatrix} \right\| = 0 \tag{45}$$

$$(n)\ \text{self-}d\quad \left\| \begin{bmatrix} \lambda & -1 \\ 1 & \lambda \end{bmatrix} + \begin{bmatrix} (\lambda+1)\frac{1}{\mu} & 0 \\ 0 & (\lambda+1)\frac{1}{\mu} \end{bmatrix} \right\| = 0 \tag{46}$$

$$(n)\ \text{mutual-}d\quad \left\| \begin{bmatrix} 1 & \lambda \\ -\lambda & 1 \end{bmatrix} + \begin{bmatrix} (\lambda+1)\frac{1}{\mu} & 0 \\ 0 & (\lambda+1)\frac{1}{\mu} \end{bmatrix} \right\| = 0 \tag{47}$$

$$(n+1)\ \text{mixed-}d\quad \left\| \begin{bmatrix} 1 & 0 & 1 \\ 0 & \lambda & -1 \\ -\lambda & 1 & 0 \end{bmatrix} + \begin{bmatrix} (2)\frac{1}{\mu} & 0 & 0 \\ 0 & (\lambda+1)\frac{1}{\mu} & 0 \\ 0 & 0 & (\lambda+1)\frac{1}{\mu} \end{bmatrix} \right\| = 0 \tag{48}$$

Thus, the actual λ-roots of the programmed systems, as indicated in the previous section, may be obtained from the determinantal equations 43 through 48. These include not only the synthesis matrices but also the respective error matrices for each of the equivalent systems.

Table I. Computed and Measured λ-roots of Differential Equation $(1+d^2)x=0$ Programmed According to Equations 34 Through 39

Synthesis System	Computed Roots*	Experimentally Measured Roots
GEDA	no computed values	$-0.17\times10^{-3}+i\,1.00$
		$-0.17\times10^{-3}-i\,1.00$
$(n+1)$ self-d	$-0.98\times10^{-3}+i\,1.00$	$-1.07\times10^{-3}+i\,1.00$
	$-0.98\times10^{-3}-i\,1.00$	$-1.07\times10^{-3}-i\,1.00$
	-0.50×10^{5}	
	-1.00×10^{5}	
	-1.00×10^{5}	
$(n+1)$ mutual-d	$-1.02\times10^{-3}+i\,1.00$	
	$-1.02\times10^{-3}-i\,1.00$	unstable
	-1.00×10^{5}	
	$2.49\times10^{4}+i\,6.62\times10^{4}$	
	$2.49\times10^{4}-i\,6.62\times10^{4}$	
(n) self-d	$-0.99\times10^{-3}+i\,1.00$	$-1.06\times10^{-3}+i\,0.99$
	$-0.99\times10^{-3}-i\,1.00$	$-1.06\times10^{-3}-i\,0.99$
	-1.00×10^{5}	
	-1.00×10^{5}	
(n) mutual-d	$-1.01\times10^{-3}+i\,1.00$	$-1.10\times10^{-3}+i1.01$
	$-1.01\times10^{-3}-i\,1.00$	$-1.10\times10^{-3}-i\,1.01$
	$-1.01\times10^{3}+i\,1.00\times10^{5}$	
	$-1.01\times10^{3}-i\,1.00\times10^{5}$	
$(n+1)$ mixed-d	$-1.00\times10^{-3}+i\,1.00$	$-1.19\times10^{-3}+i\,1.00$
	$-1.00\times10^{-3}-i\,1.00$	$-1.19\times10^{-3}-i\,1.00$
	-1.00×10^{3}	
	$-2.50\times10^{2}+i\,6.62\times10^{2}$	
	$-2.50\times10^{2}-i\,6.62\times10^{2}$	

* Amplifier characteristics of the experimental laboratory computer were designed and measured such that $k=1{,}000$ and $\tau=0.01$, except for case $(n+1)$ mixed-d which had $k=1{,}000$ and $\tau=1.0$.

It is helpful at this stage to consider actual numerical results, particularly for a comparison of the theory with physical systems. Accordingly the coefficients in the matrices 34 through 39, and 43 through 48, were taken to represent micromhos and microfarads. The λ-roots of the determinantal equations 44 through 48 were then computed and are listed in Table I. No computations were made for equation 43 since the amplifier characteristics μ_a, μ_b, and μ_c were unknown.

The computed λ-roots were then compared with physically measured λ-roots obtained with a laboratory-type electronic analogue computer. This computer was designed with carefully controlled amplifier characteristics, the $\mu_i(d)$, expressible over the range of interest of the variables by the equation

$$\pm W=\mu V \tag{49}$$

$$\mu(d)=\frac{k}{\tau d+1} \tag{50}$$

Equation 50 may be recognized as representing a linear amplifier with a gain k, a single resistance-capacitance $=\tau$ cutoff, a single input V, and bipolar outputs $+W$ and $-W$. The same $\mu(d)$ characteristics were of course assigned in the computations of the λ-roots from equations 44 through 48.

The experimentally measured λ-roots determined from the oscillographic records obtained with the laboratory computer are also listed in Table I. As may be noted, the agreement between computed and measured λ-roots is well within experimental errors, particularly in view of the 10 per cent tolerance of all dinode elements (conductances and capacitances) utilized in the experimental computer. Included in this table are measured λ-roots obtained with a *GEDA L3* computer (Goodyear Aircraft Corporation) programmed according to equation 34.

The initial portions of the oscillograms of the solutions of equation 33 taken with the *GEDA* and the laboratory computers are shown in Fig. 10. The oscillograms show solutions (for identical initial conditions) according to the programming equations 34 through 39, and ideographs Figs. 4 through 9 respectively. Clearly, all the systems yield substantially identical solutions to equation 33, with one exception. This is the $(n+1)$ mutual-d synthesis, equation 36, which as predicted by the determinantal expression 45 turns out to be an unstable electronic network. This is confirmed physically by the corresponding oscillographic record for this case in Fig. 10.

It may be noted from the oscillographic records that the decay of the amplitude of the solution obtained with the laboratory computer (a sensitive test) compares not unfavorably with the solution obtained with the *GEDA* computer, although the experimental computer gain $k=1{,}000$ is far less than the quoted gain of the *GEDA*. Finally, it may be inferred from the oscillograms that the error matrices have but a negligible effect upon the magnitude of the coefficients of the solution, although all-important in so far as the electronic network stability is concerned.

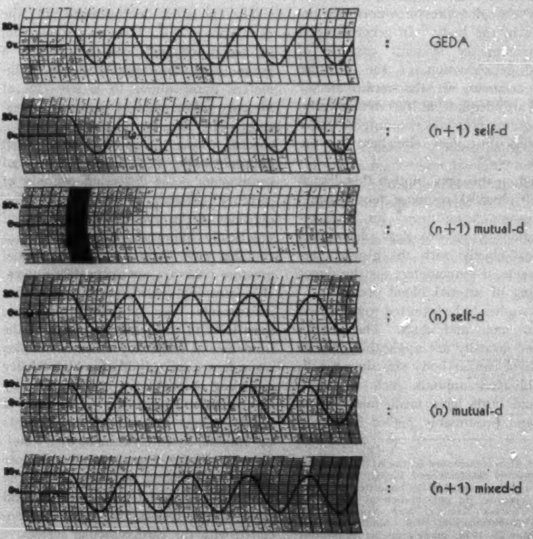


Fig. 10. Oscillograms of solutions of $(1+d^2)x=0$ displayed by GEDA and experimental computer. Sensitivity 2 volts per millimeter; chart speed 5 millimeters per second

Conclusions

The application of matrix methods provides a rational and powerful method of analysis of electronic analogue computer networks. The matrix method is completely general, and may be applied not only to the study of computer stability but also to other problems such as the actual evaluation of the magnitude of the coefficients of the solution as well as the determination of the λ-roots of the electronic network; the study of the best amplifier characteristics for computer applications; the effects of precision of network components, and of stray parameters. The latter is not unimportant, particularly for matrices with zero entries on their principal diagonal, such as in the synthesis matrix in equation 39.

Other attributes of the matrix method may be found in reference 1.

Reference

1. Matrices in Analogue Mathematical Machines, Pierre M. Honnell, Robert E. Horn. *Journal*, Franklin Institute, Philadelphia, Pa., vol. 260, no. 3, Sept. 1955, pp. 193–207.

Physical Response Requirements of Pressure Transducers for the Reproduction of Physiological Phenomena

E. H. WOOD
NONMEMBER AIEE

THE pressure variations generated by the heart beat are perhaps the most rapidly changing pressures normally occurring in the body. In general therefore manometer systems capable of faithful reproduction of the pressure pulses occurring in the cardiovascular system are adequate, at least dynamically, for recording pressures encountered under usual circumstances elsewhere in the body.

Based on this presumption this discussion of physical response requirements of pressure transducers for the reproduction of physiological phenomena will deal chiefly with the physical requirements of manometers for the direct recording of arterial blood pressure.

The studies to be reported were carried out on men and dogs. The results obtained actually are applicable only to animals of similar body size since small, warm-blooded animals such as mice, rats, and birds have much faster heart rates and presumably higher frequency variations in blood pressure than those encountered in man.

The pressures encountered in the circulation and the body as a whole will, in nearly all instances, fall within a range from approximately 300 millimeters (mm) above to a few mm of mercury below ambient barometric pressure. The sensitivity required in recording these pressures varies from approximately 1 mm up to several centimeters' (cm) deflection per mm of mercury pressure. The highest frequencies of the dynamic components of practically important magnitudes in a complex wave form such as an arterial pressure pulse are not accurately known. It is generally considered, however, that instruments with a uniform dynamic sensitivity to the tenth harmonic of the fundamental frequency of such complex wave forms are suitable for high-fidelity recording of the wave concerned.[1] By this criterion, since the heart rate of human beings seldom exceeds 240 beats per minute, as an instrument with a uniform sensitivity from 0 to 40 cycles per second (cps) should be adequate for the recording of arterial blood pressure and most other physiological variables associated with the cardiovascular system.

It should be emphasized that certain requirements of manometer systems are equally or more important than are the dynamic characteristics. These are: high stability, linear calibration, usability with long leads, insensitivity to movement, temperature humidity and acceleration, imperviousness to electrolyte solutions, simplicity of operation, and construction for ease of sterilization and removal of air bubbles entrapped in the hydraulic system. Of these requirements high stability and simplicity of operation are, from a practical viewpoint, of greatest importance. It is of little value for a manometer to have extremely good dynamic characteristics, allowing reproduction of the fastest dynamic components of the pressure pulses, if instability of its base line or of its sensitivity results in errors in recorded pressures of greater magnitude than the relatively small dynamic components of the pressure wave which it is capable of reproducing.

Although it is generally agreed that a manometer system with uniform sensitivity from 0 to 50 cps will accurately reproduce arterial pulse waves,[2] there are no direct data in the literature concerning how much this dynamic response can be decreased without producing significant distortion of the recorded pulse wave. This is a practically very important question, especially in reference to direct blood pressure studies in human beings. It determines, for any given manometer system, the minimal dimensions of hypodermic needles or catheters that must be introduced into the vascular system in order to permit accurate recording of pressures therefrom. The desirability of using the smallest possible needles and catheters when studies are carried out on unanesthetized human beings is obvious.

The remainder of this presentation will be devoted to this problem of the minimal dynamic response characteristics of manometer systems required for adequate recording of vascular pressure pulses in man.[3]

Paper 56-81, recommended by the AIEE Electrical Techniques in Medicine and Biology Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted May 25, 1953; made available for printing December 1, 1955.

E. H. Wood is with the University of Minnesota and the Mayo Clinic, Rochester, Minn.

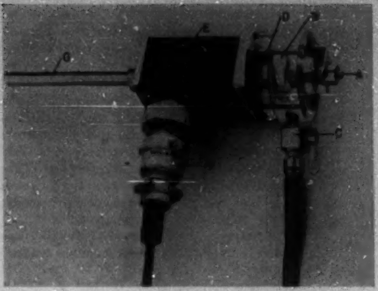


Fig. 1. Strain-gauge manometer, Statham model P6-15G-250, adapted for physiological studies

By permission of Science, Washington, D. C.

Fig. 3. Apparatus for testing dynamic response of manometer systems by subjecting them to square-wave pressure change[9]

The perfection of electric manometers for recording pressures simplifies the collection of direct data on this point since it is a simple matter to change the over-all dynamic response of such manometers rapidly and reproducibly. Strain-gauge[4] capacitance[5] and variable reluctance[6,7] pressure transducers were used in these studies. Of these the unbonded strain-gauge manometers, because of their stability and simplicity of operation, proved the most useful.

Such a manometer is illustrated in Fig. 1. The hypodermic Luer-lok fitting for attachment of needles or catheters is fitted onto a lucite housing specially constructed for ease of detection and removal of entrapped air bubbles. The stopcock allows flushing of the manometer chamber and attachments with sterile saline solution. The electric connections consist of two pairs of leads, one pair of which connects to a source of direct current for maintaining a constant voltage across the strain-gauge wheatstone bridge circuit and the other connects to a high-performance oscillographic galvanometer which records the output of the strain gauge without electronic amplification.

Fig. 2. Experimental setup used for studying the minimal dynamic response characteristics of manometer systems necessary for recording peripheral arterial pressure pulses in man

At *A* in Fig. 1, is shown the Luer-lok fitting for attachment of hypodermic needles, venous or arterial catheters and the like. At *B* is the lucite chamber filled with sterile saline solution containing an anticoagulant (heparin, 40 milligram per liter); this solution can be flushed through the attached needle or catheter by opening stopcock *C* to a pressurized wash bottle reservoir containing the solution. At *D* is the lucite chamber filled with oil and containing the pressure-sensing diaphragm of the manometer; chambers *D* and *B* are separated by a passive diaphragm of polythene, 0.006 inch thick. At *E* is the housing of the Statham strain-gauge unit; At *F* is the electric connector to the indicating-galvanometer circuit and d-c source. At *G* is the rod for mounting the manometer. The lucite chamber *B* is dome-shaped and polished for ease of visualization and the removal of entrapped air bubbles.

A manometer assembly commonly used in laboratories for continuous recording of arterial pressure[8,9] consists of a 20-gauge needle connected to a glass hypodermic adapter, thence to a 25-cm length of polythene tubing (0.085-inch inside diameter (ID), 0.135-inch outside diameter), connected to the strain gauge via a 2-way stopcock; see Fig. 2. The stopcock makes possible withdrawal of arterial blood samples from the same needle through which arterial pressure is recorded.[8]

In Fig. 2 the two strain-gauge manometer systems are connected to indwelling needles in the radial and brachial arteries. The needles are kept free from clot formation by frequent flushing with heparinized saline solution contained in the reservoir wash bottle system which is pressurized by the hand-bulb Tycos manometer assembly. The zero reference point for the manometers is established by opening them to the wash bottle when its meniscus is at midchest level. The manometers are calibrated *in situ* and simultaneously in a similar fashion by applying known pressures above the fluid level in the wash bottle. One manometer system is used to monitor the arterial pressure pulses continuously when the effects of variation of dynamic response of the second manometer on the pressure pulse recorded by that manometer are under study.

The dynamic response characteristics of such a manometer assembly are rather difficult to predict since it cannot be considered certain that the well-established frequency-damping relationships which apply to a simple instrument with one degree of freedom would be applicable to a heterogeneous assembly such as this or

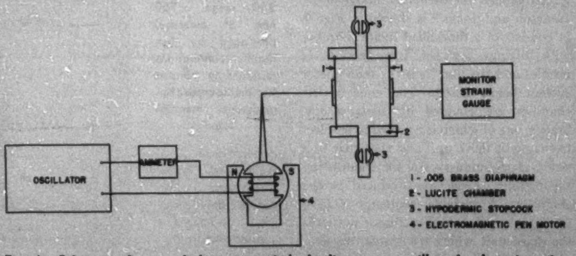


Fig. 4. Schematic diagram of electromagnetic hydraulic-pressure oscillator for dynamic testing of manometer systems

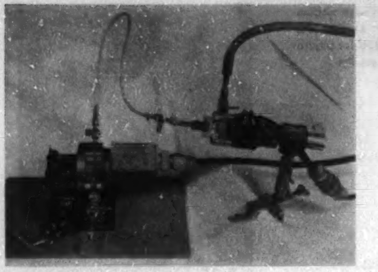


Fig. 5. Pressure changer assembly of electromagnetic hydraulic pressure oscillator

many of the other more or less complicated manometer assemblies commonly used in physiological studies.

Owing to this, and to the large variations in dynamic response caused by even minute air bubbles entrapped in the connections of the hydraulic system, it is important to have a convenient method of studying the dynamic response of such manometer systems and for checking this response at the time of their use. This can be done by recording the response of the system to a square-wave pressure change as illustrated in Fig. 3.

The needle of the manometer assembly is inserted into the fluid contained in a pressure chamber. The pressure in the air space over the fluid is increased by means of a hand bulb to just under the bursting pressure of the finger cot covering the pressure chamber. The balloon is then exploded with a blow torch, thus producing an instantaneous decrease to ambient pressure.[9]

However, if the manometer does not behave as a simple instrument it may not be possible to determine its dynamic characteristics from its response to a square-wave pressure change alone. Hydraulic pressure oscillators have been specially designed to study the response of such manometer systems to equal amplitude sine-wave pressure variations of variable frequency. A schematic diagram of such an oscillator designed by Isaacson and Jones[10] is shown in Fig. 4. It consists of a fluid-filled lucite chamber with fittings for the introduction of needles and the like from the manometer system being tested. The ends of the chamber are covered by metal membranes, one of which is coupled to an electromagnetic drive unit and the other to a strain-gauge manometer for monitoring the pressure variations produced in the chamber. A variable frequency (0.5 to 50,000 cps) electronic oscillator actuates the drive unit which translates the sine-wave oscillator current into sine-wave pressure variations in the chamber. Constant-pressure amplitude at varying frequencies is achieved by maintenance of a constant current to the drive unit, monitored by a milliammeter in the drive unit circuit. Square-wave pressure variations can also be produced by making and breaking a source of direct current to the electromagnetic driver.

Such an oscillator in use is shown in Fig. 5.

The needle of the manometer assembly undergoing test is inserted into fluid-filled lucite placed in the pressure chamber. The electromagnetic drive unit is to the left of the pressure chamber, and the strain-gauge for monitoring the generated pressure changes is on right.

The alteration in dynamic response of a strain-gauge manometer by the various components of the multipurpose manometer system shown previously is illustrated in Fig. 6.

The upper panels show the response of the water-filled gauge alone; the top trace is from the monitor gauge and the bottom from the gauge being tested. The first panel represents the square-wave response and subsequent panels show the response to sine waves of increasing frequencies. The gauge alone had a sharp resonant peak at 190 cps. Attachment of a 5-cm 20-gauge needle (0.6-mm ID) to the gauge reduced the resonant frequency to 60 cps. The polythene tube assembly had a similar effect. Combination of polythene tube and needle to complete the hydraulic system reduced the resonant frequency to 40 cps. Additional electromagnetic damping of the oscillographic galvanometer produced a reasonably satisfactory manometer system with a uniform sensitivity from 0 to 40 cps. The time lag in this manometer system of about 10 milliseconds was practically independent of the frequency in the range of from 0 to 40 cps; Fig. 6 (B). In Fig. 6(A) the top tracing in each panel is the recording from the gauge monitoring the pressure chamber and the bottom tracing is the recording from the manometer system being tested. The 20-gauge hypodermic needle was 5 cm in length and 0.056 cm in ID. The polythene tubing assembly consisted of a glass hypodermic adapter for attachment to the needle, a 26-cm length of polythene tubing (2.1-mm ID, 3.4-mm outside diameter), and a 2-

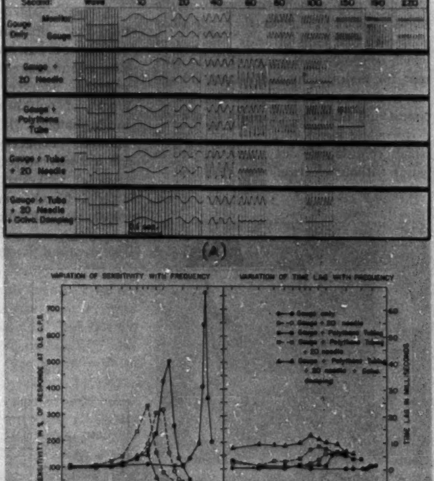


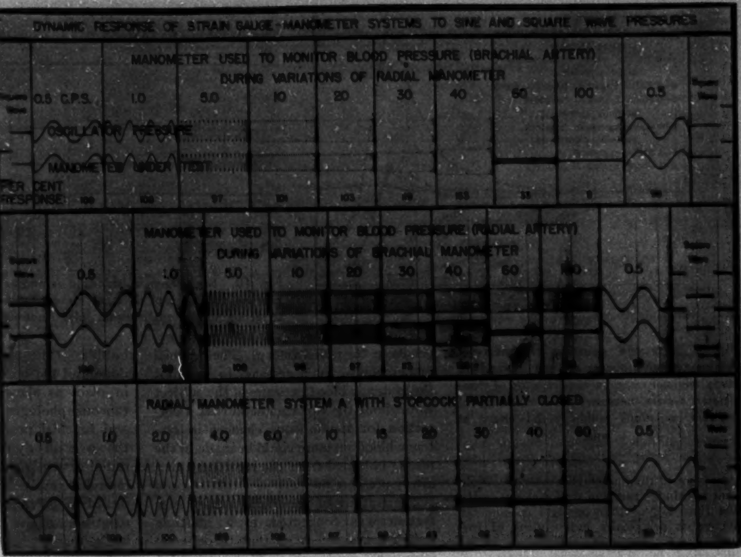
Fig. 6. Alteration of dynamic response of strain-gauge manometer (Statham gauge, P6-15G-250, range ±760 mm of mercury) produced by hydraulic connections required to convert it to multipurpose hypodermic manometer[9]

A—Response of water-filled gauge
B—Dynamic response data measured from recording illustrated in A

By permission of Science, Washington, D. C.

Fig. 7. Dynamic response of three different strain-gauge manometer systems to sine- and square-wave pressure variations

By permission of Science, Washington, D. C.



way hypodermic stopcock to permit blood sampling and connection to the manometer proper. Galvanometer damping (lower panel) consisted of the use of a galvanometer with a resonant frequency of 40 cps utilizing an external damping resistance of 70 ohms (optimal damping resistance, 250 ohms).[9]

In Fig. 6(B) note that the gauge alone, Fig. 1, has a high degree of overshoot at its resonant frequency of 190 cps. However, each hydraulic connection added to complete the system reduces this resonance point so that the complete assembly with needle attached has a resonant frequency of about 40 cps and a damping coefficient of approximately 0.2 of critical. The damping of the complete system was increased to a more optimal range by utilizing a recording galvanometer, the physical characteristics and damping of which were selected to compensate for the

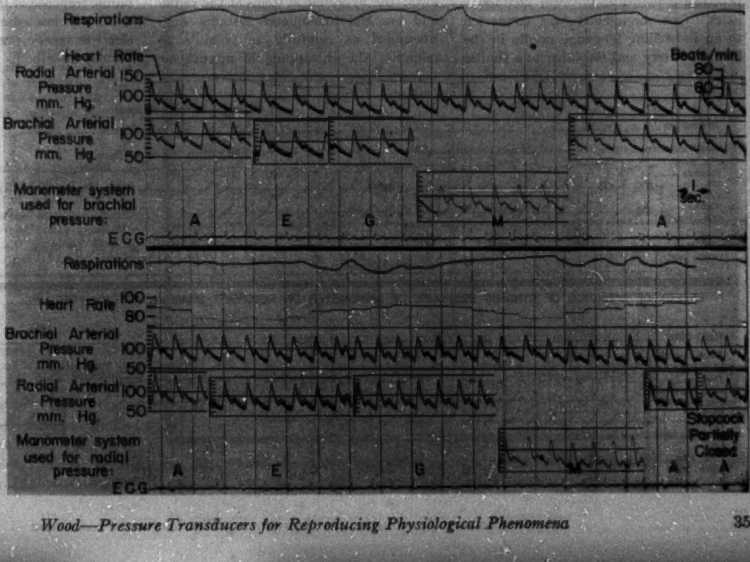


Fig. 8. Effect of variation in dynamic response of recording manometer system on contours and pressures of brachial and radial arterial pressure pulses recorded from normal subject, J. H.

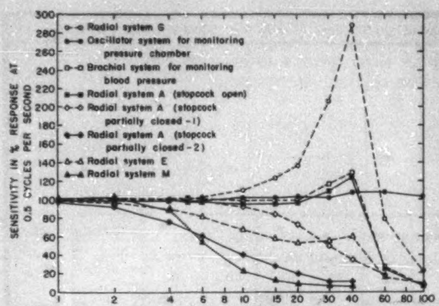


Fig. 9 Average dynamic responses of manometer-galvanometer systems utilized to study minimal response requirements of manometer system necessary for recording of pressure variations within radial artery at wrist

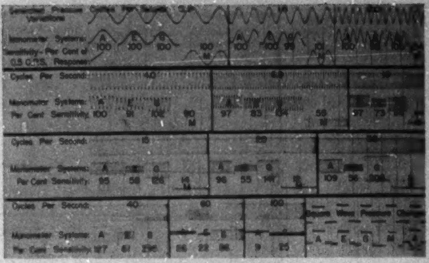


Fig. 10. Recordings of dynamic responses to sine- and square-wave pressure variations of different manometer-galvanometer systems used to record radial arterial pressure pulse of normal subject J. H., see Fig. 8

overshoot of the manometer system. Note that within its range of uniform sensitivity the time lag of the complete manometer system is practically independent of frequency, a condition which minimizes distortion in the recorded contour of complex wave forms such as arterial pressure pulses.

The experimental setup for studying the minimal dynamic response characteristics needed for recording of peripheral arterial pressure pulses in man is illustrated in Fig. 2. Two strain-gauge manometer systems were used with practically identical physical characteristics. One manometer was connected to an indwelling 20-gauge needle in the brachial artery and the other to a similar needle in the radial artery. The brachial manometer was necessary to monitor the systemic blood pressure of the subject throughout the experiment so that a correction for spontaneous changes in systemic blood pressure could be made in the results obtained from the recorded radial artery pressure pulses.

The output of the radial manometer system was connected to a multiple switch arrangement so that the electric variations produced could be recorded interchangeably by four different galvanometers of widely different dynamic response. Also the mechanical damping and resonant frequency characteristics of the manometer proper could be varied by partially closing the stopcock between the manometer and the arterial needle.

The experiment was performed as follows: the hydraulic systems of the radial and brachial manometer systems were assembled as carefully as possible to minimize the entrapping of minute air bubbles. The arterial punctures were done and the needles were lightly taped in place as shown. Continuous simultaneous photographic records of radial and brachial pressure were taken while the over-all dynamic response of the radial system was varied by switching it interchangeably to the four different recording galvanometers. After several repetitions of this procedure the stopcock in the radial system was carefully turned until an appreciable difference was thought to be detected in the galvanometer trace being recorded. In part of the experiments the procedure was repeated, the dynamic response of the brachial manometer system being varied while the radial system was kept constant.

When the recordings were completed the needles were withdrawn and the dynamic response of the manometer systems used was immediately recorded by the electromagnetic-pressure oscillator previously described.

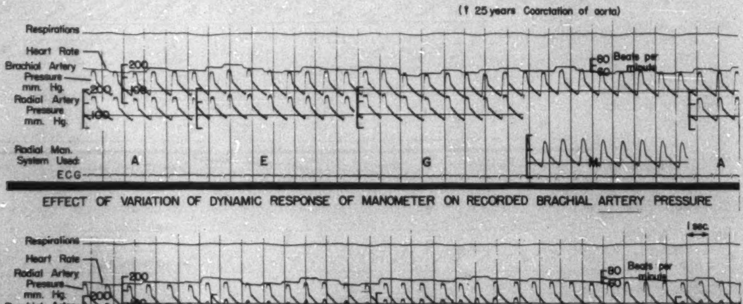

(♀ 25 years Coarctation of aorta)
Respirations
Heart Rate
Brachial Artery Pressure mm. Hg.
200
100
Radial Artery Pressure mm. Hg.
200
100
80
60
Beats per minute
Radial Man. System Used:
A
E
G
M
A
ECG
EFFECT OF VARIATION OF DYNAMIC RESPONSE OF MANOMETER ON RECORDED BRACHIAL ARTERY PRESSURE
1 sec.
Respirations
Heart Rate
Radial Artery Pressure mm. Hg.
200
100
Brachial Artery Pressure mm. Hg.
200
100
80
60
Beats per minute
Brachial Man. System Used:
A
E
G
M
A
ECG

Fig. 11. Effect of variation in dymanic response of recording manometer system on contours and pressures of radial and brachial arterial pressure pulses recorded from patient with coarctation of aorta

Table I. Average Effect of Variation in Dynamic Response of Manometer System of Systolic Pressure (mm Hg) Recorded from Radial Artery of Normal Subjects*

No. of subjects	11	11	6	11	11	5
Brachial systolic pressures during radial pressure recordings	118	118	123	118	119	156
Manometer system used for radial pressure recordings	*G*	*A*	A_1	*E*	*M*	A_2
Recorded radial systolic pressure	123	124	129	120	122	144
Difference (brachial-radial systolic pressure)	−5.2	−6.2	−5.8	−1.6	−3.5	+11.4
P value of difference†	<0.01	<0.001	<0.02	<0.2	<0.05	<0.02

* Average age of subjects: 31 (24 to 38) years; average heart rate: 67 (59 to 83) beats per minute.

† Differences with *P* values of less than 0.02 considered to be statistically significant. This also applies to Tables II and III.

Table II. Effect of Variation of Dynamic Response of Radial Manometer on Differences Between Simultaneously Recorded Brachial and Radial Systolic Pressures

	Radial Minus Brachial Systolic Pressures Recorded with Manometer System A Minus Differences Obtained when Radial Pressures Were Recorded with Manometer System				
	G	A_1	E	M	A_2
No. of subjects	11	6	11	11	5
Difference, mm Hg mean	−1.1	−1.3	−4.5	−3.0	−17.0
Range	−5 to 3	−7 to 3	−9 to 2	−7 to 6	−30 to −8
P value of difference	<0.3	<0.5	<0.01	<0.05	<0.02

Recordings of the dynamic response of three different manometer systems are shown in Fig. 7. The top trace in each panel is a recording of the pressures, and the bottom trace is the response of the manometer system.

The top trace on each panel records pressure variations generated in the pressure oscillator chamber; see Fig. 5. The bottom trace is the response of the manometer system under test to the following pressure variations: a square-wave pressure change; sine-wave pressure variations of 0.5, 1, 5, 10, 20, 30, 40, 60, and 100 cps. The numerals toward the top of each panel are the frequencies of the sine-wave pressure variations in cps; those at the bottom, the response of the manometer under test expressed as per cent of the response at 0.5 cps: the dynamic responses of the manometer systems determined immediately after being utilized to record brachial and radial arterial pressures on normal subject J. H.; see Fig. 8.

Note that the response of the manometer systems utilized to monitor radial and brachial arterial pressure were essentially uniform out to 20 cps. The sensitivity increased thereafter to attain a peak response at about 40 cps. Partially closing the stopcock on the radial manometer system (bottom panels) decreased the range of uniform response to 10 cps with decreasing sensitivity to frequencies above this range.

The average dynamic responses of the various manometer systems used are shown in Fig. 9. The four different manometer-galvanometer systems were assembled so that one system, designated as *G*, was badly under-damped, overshooting approximately 300% at 40 cps; the second, designated as *A*, was approximately optimally damped, the response being flat to 20 cps and overshooting 20% at 40 cps; the third system *E* was overdamped, the response cutting down to 90% at 4 cps and to 60% at 20 cps; the fourth system *M* was also overdamped with a uniform response to 4 cps, cutting down to 50% at 6 cps.

Note that the upper range of uniform response for the different systems used varies from about 40 to less than 2 cps. The average figures for manometer systems designated radial systems *G*, *A*, *E*, and *M* are based on determinations carried out after the completion of the studies on each of 16 subjects. The coefficient of variation of the response at the various frequencies studied ranged from 1 to 10%.

The response of these manometer-galvanometer systems to sine-wave pressure variations when coupled interchangeably to these four different galvanometers is shown in Fig. 10. At 0.5, 1.0 and 2.0 cps the response of the four systems *A*, *E*, *G*, and *M* was uniform; however, at 10 cps the *E* system was down to 73%, the *G* was overshooting to 114% while the *M* was down to 24%; at 30 cps the *A* system response was 109%, the *E*, 56%; the *G*, 205% and the *M*, 8% of the actual pressure variation.

In Fig. 9 the top trace on each panel is a recording of the pressure variations generated in the pressure oscillator chamber; see Fig. 5. The bottom trace is the response of the manometer system under test as its output was recorded by four different galvanometers, designated *A*, *E*, *G*, and *M*, of widely different dynamic response.

Note that in each panel each of the four manometer-galvanometer systems was subjected to identical sine-wave pressure variations at frequencies ranging from 0.5 up to 100 cps. System *A* was slightly

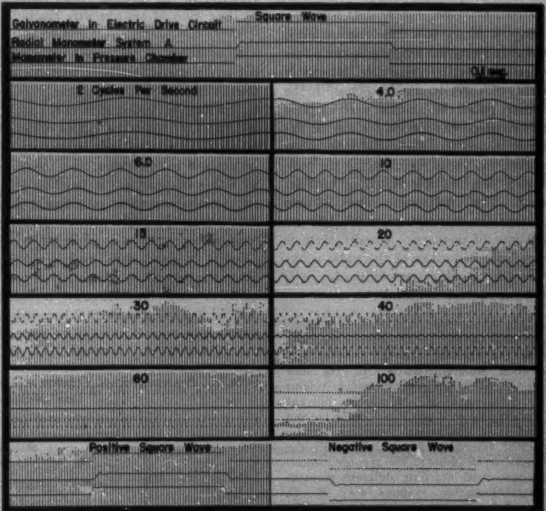


Fig. 12. Photokymographic recording of response of manometer system to square-wave and sine-wave pressure variations

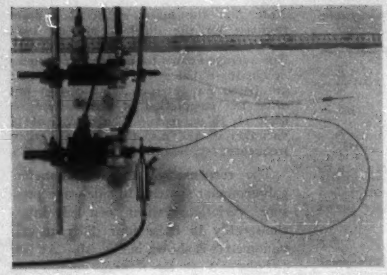


Fig. 13 (above). Manometer systems utilized for recording peripheral arterial and aortic pressure pulses in man

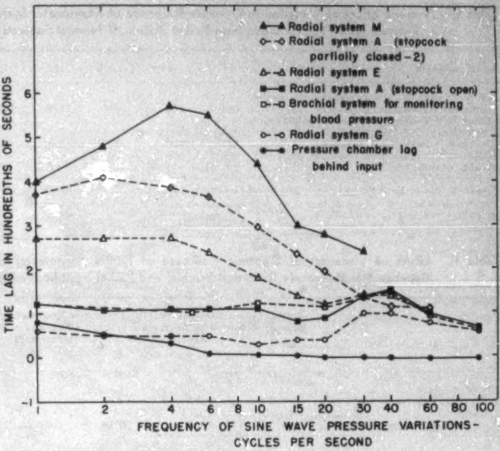


Fig. 14 (right). Average time lag to sine-wave pressure variations of manometer-galvanometer systems utilized to study minimal dynamic response requirements of manometer system necessary for recording of pressure variations in peripheral arteries of man

underdamped, showing a peak sensitivity of 127% at 40 cps. System *E* was badly overdamped, showing a decrease in sensitivity from 91% at 4 cps to 61% at 40 cps. System *G* is badly underdamped having a peak response at 295% at 40 cps. System *M* exhibits an essentially uniform response out to 4 cps with a sharp cutoff in sensitivity to frequencies above this range. The response of these systems to square-wave pressure variations is shown in the lower right-hand panel. The thin vertical black lines delineate 1-second intervals.

Fig. 8 shows the effect on recorded radial and brachial artery pressure pulses when recorded by these four systems of widely different dynamic response. In the bottom panel the brachial arterial pressure recorded continuously while the dynamic response of the radial manometer system was varied by changing the position of the multiple-switching arrangement. The horizontal lines drawn through the tracings are at 50, 100 and 150 mm of mercury. Visually there is very little distortion of the pulse wave, even when the *M* system with only a 50% response at 6 cps was used.

In Fig. 8 the top panel is a continuous recording of the radial arterial pressure while the brachial arterial pressure was recorded simultaneously with the four different manometer-galvanometer systems (*A*, *E*, *G*, and *M*) of widely different dynamic response; see Fig. 9. The bottom panel is a continuous recording of the brachial arterial pressure (top) while the radial arterial pressure was recorded simultaneously with the four different manometer-galvanometer systems (*A*, *E*, *G*, *M*, and *A* with a stopcock in the hydraulic connections partially closed). The tracings of the dynamic responses of these manometer systems, exactly as used for these pressure pulse recordings, are shown in Figs. 7 and 10. The dynamic response tracings, Figs. 7 and 10, were recorded without disassembly of any portion of the manometer systems immediately after withdrawal of the needles (still connected to the manometer systems) from the radial and brachial arteries.

Note that in spite of the wide variation in the dynamic response of the manometer-galvanometer systems the contours and pressures of the pressure pulses recorded by these different systems were practically identical.

Experiments of this type were carried out on 11 normal subjects and five patients with aortic pathologic changes (aortic insufficiency, coarctation of the aorta, see Fig. 11). Only the effects obtained by varying the radial arterial manometer system will be considered, since the studies with variations in the brachial arterial manometer systems gave closely similar results. The differences in pressure recorded from the radial artery when the different systems were used could be demonstrated only by comparing all pressures to the simultaneously recorded brachial arterial pressure. When such a comparison is made, it must be realized that radial systolic pressure uniformly exceeds the systolic pressure recorded simultaneously in the brachial artery.[11]

In Fig. 11 the top panel shows the effects of variations in the manometer system on the recorded radial pulse using the simultaneously recorded brachial arterial pressure as the basis of comparison. The bottom tracing shows a similar study of the recorded brachial arterial pulse using the radial arterial pressure as the basis of comparison. The dynamic response of the four different manometer-galvanometer systems *A*, *E*, *G*, *M* is shown in Fig. 9. Note that there were no significant differences in the systolic or diastolic pressures recorded by the four manometer-galvanometer systems. However, the true contour of the peak of the pressure wave was not accurately reproduced by system *M* (the dynamic response of which was uniform out to only 4 cycles per second). Otherwise there were no evident differences in the contours of the pulses recorded by the four systems.

Average radial systolic pressures along with simultaneously recorded brachial arterial systolic pressures are shown in Table I. When manometer-galvanometer system *G* was used, the brachial minus radial systolic pressure difference

Table III. Average Effect of Variation of Dynamic Response of Manometer System on Build-up Time* of Recorded Radial Pulses

	Differences Between Build-up Times* of Radial Pulses Recorded with Manometer System A and Manometer System				
	G	A_1	E	M	A_2
No. of subjects	11	6	11	11	5
Mean difference, seconds	0.001	−0.002	−0.013	−0.034	−0.020
Range	−0.02 to 0.01	−0.01 to 0.01	−0.03 to 0.00	−0.05 to 0.02	−0.03 to 0.01
P value of difference	<0.80	<0.70	<0.001	<0.001	<0.01

* Elapsed time from onset to systolic peak of pulse wave.

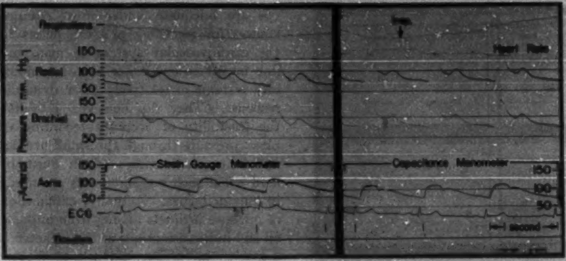


Fig. 15. Comparison of aortic pressure pulses recorded in immediate sequence from same subject by strain-gauge and by capacitance manometer-catheter systems

was −5 mm of mercury; with the *A* system this difference was −6; with the *E* system, −2; and with the *M* system, −3.5 mm of mercury. To demonstrate whether or not these small differences in pressure were significant, it was necessary to make a paired comparison of the the brachial-minus-radial pressure differences in each subject. The pressure difference recorded was used with the *A* system as the standard. Table II shows the differences in the radial minus brachial systolic pressures recorded with manometer system *A* minus the differences obtained with the other systems. These differences for manometer-galvanometer system *G* averaged 1 mm of mercury and were not statistically significant; for the *E* system, a statistically significant reduction of 4.5 mm of mercury in the radial systolic pressure was obtained, while a questionably significant reduction of 3 mm of mercury was obtained for the *M* system. The results when damping of the radial manometer was produced by partially closing the stopcock were divided into two groups A_1 and A_2 on the basis of the dynamic responses; see Fig. 9. The A_1 system included experiments in which the dynamic response of the system was more than 90% at 6 cps and less than 50% at 40 cps. The response of the A_2 group was less than 90% at 4 cps. Radial systolic pressure was not significantly affected when recorded with the A_1 system (flat to 6 cps). The A_2 system, however, with an average response of 80% at 4 cps, reduced the radial pressure on the average 17 mm of mercury.

The diastolic and mean radial pressures were not affected by any of the manometer systems used. Results similar to those stated previously were obtained from five patients with aortic pathologic changes (aortic insufficiency or coarctation; see Fig. 11).

From these studies it is concluded that, under the conditions of these experiments on resting normal and abnormal subjects with heart rates varying from 60 to 108 beats per minute, a manometer system with a uniform dynamic response out to 6 cps was adequate for accurate recording of the systolic, diastolic, and mean radial or brachial arterial pressure. Actually, practically significant errors in systolic pressure could not be demonstrated until the range of response was reduced to less than 4 cps.

An analysis has been made of the distortion of the time characteristics, i.e., the contour of the arterial pulse associated with variations in the dynamic response of the manometer system. The time delay characteristics of the various manometer-galvanometer systems were determined by recording their responses to sine-wave and square-pressure variations at a paper speed of 245 mm per second; see Fig. 12.

The recordings in Fig. 12 were taken at a paper speed of 245 mm per second to determine the time delay characteristics of a hypodermic manometer system. The top trace is a recording of the current variations in the circuit actuating the electromagnetic drive unit of the pressure oscillator; see Fig. 4. The bottom trace is a recording of the pressure variations generated in the chamber of the oscillator; see Fig. 5. The middle trace is a recording of the response of the manometer system under test, Fig. 13, to these pressure variations.

Note that there was practically no measurable time lag between the input driving current to the oscillator and the pressure variations generated therefrom. The delay in the test manometer system to sine-wave pressure variations was about 0.01 second and practically independent of the frequency.

The average time lag of the different manometer-galvanometer systems is shown in Fig. 14. The time lag of the various systems varied from 5 to 50 milliseconds. With exception of system *M* the time lag of the individual systems tended to be independent of frequency within the frequency range for which the amplitude response of the system was uniform, see Fig. 9. Table III shows differences between the pulse build-up times recorded with manometer system *A* and the other systems used. This time interval from the onset to the peak of the radial pulse wave was not significantly

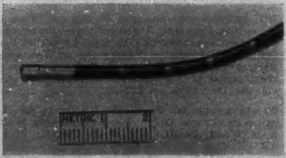
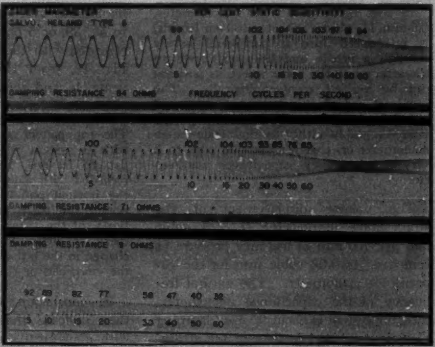


Fig. 16. Miniature variable reluctance intracardiac manometer attached to 8F Coumand cardiac catheter

Fig. 17. Dynamic response of miniature intracardiac manometer to equal amplitude pressure variations generated by hydraulic pressure oscillator. Three damping conditions illustrated were employed to record pressure tracings shown in Fig. 18



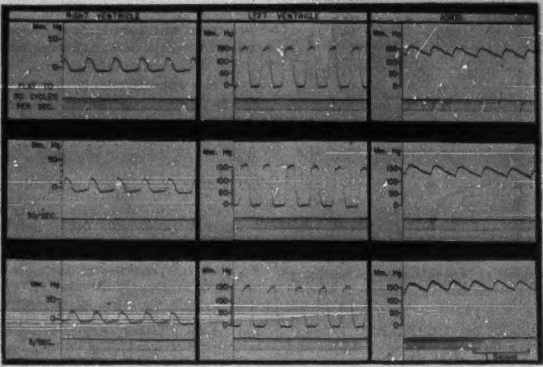


Fig. 18. Effect of rapid interchange of manometric systems with different characteristics of dynamic response on contours of pulse waves recorded from right ventricle, left ventricle, and arch of aorta in an intact dog

different when the *G*, *A*, or A_1 systems were used. A significant prolongation of this interval was detected, however, in the systems with a uniform dynamic response out to less than 4 cps.

These results, indicating that manometers with uniform dynamic sensitivity out to 6 to 10 cps are adequate to record peripheral arterial pulses, possibly may not apply to central arterial pulses or the pressure pulses generated in the chambers of the heart proper. Data have been collected which indicate, however, that such manometer systems are generally adequate for this purpose.

Fig. 13 shows a manometer assembly used for recording pressure pulses from the aorta of man. The arterial catheter, which consists of a specially prepared polyvinyl radiopaque plastic, was developed by Peterson and associates.[5]

In Fig. 13 the 1-mm outside diameter, 0.5-mm ID plastic catheter (80-cm length) is introduced into a peripheral artery through a needle and then advanced up the artery into the aorta where its position can be verified roentgenologically. The proximal end of this catheter is attached to a 2-way stopcock so it can be connected interchangeably either to the strain-gauge manometer or, through the side arm, to a very-high-frequency capacitance manometer.

The volume displacement of the strain-gauge manometer is 0.25 cubic mm per 100 mm of mercury pressure as compared to 0.02 cubic mm for the capacitance manometer. The natural frequency of the capacitance manometer-catheter system is about 90 cps (damping coefficient 0.1) while that of the strain-gauge system is 15 cps (damping coefficient 0.4).

Fig. 15 shows pressure pulses recorded interchangeably by these two manometer systems from the aorta of a normal subject. Note that there is no apparent difference in the aortic pressure pulse contours recorded by the two systems in spite of the wide divergence in dynamic response. This was a uniform finding in studies carried out in 11 normal subjects.[12]

Somewhat similar results were obtained using a miniature variable reluctance manometer, Fig. 16, small enough to be introduced directly into large vessels or the chambers of the heart.[6,7] The output of this manometer was recorded via a carrier wave amplifier by a galvanometer, the electromagnetic damping of which could be varied interchangeably so that the over-all dynamic response of the manometer system was, Fig. 17, uniform to 50, 30, or 5 cps. The pressure pulses recorded with these manometer systems, Fig. 18, from the right ventricle, left ventricle, and aorta of dogs were not appreciably different.

The top panels of Fig. 18 were recorded with a system with a dynamic response flat to 50 cps, the middle panels were recorded with a system flat to 30 cps and the bottom panels were recorded with a system flat to 5 cps. Although the sensitivity of the three systems varies slightly there is little apparent or measurable change in the pulse contours obtained from the ventricles or the aorta.

In conclusion, these results, though they indicate that manometer systems with a uniform response out to 6 to 10 cps are adequate under most circumstances for recording pressure pulses in the cardiovascular system of man, are not meant to encourage use of manometer systems of relatively poor dynamic response. They do, however, establish the feasibility of accurate recording of blood pressure through the small needles and long plastic tubes which must be used when such measurements are made in human beings. Actually, elimination of pressure artifacts arising from impacts on and induced motions of the catheter is a much more serious obstacle to obtaining high-fidelity intracardiac pressure tracings,[9] than are the dynamic response characteristics of the manometer-catheter systems commonly used.

References

1. The Pressure Pulses in the Cardiovascular System (book), **C. J. Wiggers.** Longmans, Green and Co., Ltd., London, England, 1928, pp. 1–200.

2. Pressure Measurement in the Human Organism, **A. T. Hansen.** *Acta Physiologica Scandinavica*, Stockholm, Sweden, vol. 19 (supplement 68), 1949, pp. 1–230.

3. Study of Minimal Dynamic Response Characteristics of Manometer Systems Required for Adequate Recording of Peripheral Arterial Pressure Pulses in Man, **E. H. Wood.** *American Journal of Physiology*, Washington, D. C., vol. 163, Dec. 1950, p. 762.

4. Strain Gages: Resistance Wire (book), **E. H. Lambert.** In *Medical Physics:* The Year Book Publishers, Inc., Chicago, Ill., vol. 2, 1950, pp. 1090–98.

5. A Method for Recording the Arterial Pressure Pulse and Blood Pressure in Man, **L. H. Peterson, R. D. Dripps, G. C. Risman.** *American Heart Journal*, St. Louis, Mo., vol. 37, Apr. 15, 1949, pp. 771–82.

6. A Miniature Pressure-Recording Device, **O. H. Gauer, Erich Gienapp.** *Science*, Washington, D. C., vol. 112, new series, Oct. 6, 1950, pp. 404–05.

7. An Intracardiac Manometer: Its Evaluation and Application, **E. J. Ellis, O. H. Gauer, E. H. Wood.** *Circulation*, American Heart Association, New York, vol. 3, Mar. 1951, pp. 390–98.

8. A Direct Reading Cuvette Oximeter and Strain Gage Manometer Assembly for Measurement of Blood Oxygen Saturation and Continuous Recording of Arterial Pressure and Time-Concentration Curves of Evans Blue Dye in Arterial Blood of Man, **E. H. Wood, J. W. Nicholson, III.** *American Journal of Physiology*, Washington, D. C., vol. 163, Dec. 1950, pp. 762–63.

9. Special Instrumentation Problems Encountered in Physiological Research Concerning the Heart and Circulation in Man, **E. H. Wood.** *Science*, Washington, D. C., vol. 112, new series, Dec. 15, 1950, pp. 707–15.

10. An Electromagnetic Hydraulic Transducer for Studying Response Characteristics of Manometer Systems to Square Wave and Sine Wave Pressure Variations of Variable Frequency, **James Isaacson, R. E. Jones.** *American Journal of Physiology*, Washington, D. C., vol. 163, Dec. 1950, pp. 722–23.

11. Intraluminal Pressures Recorded Simultaneously From Different Arteries in Man, **E. H. Wood, Josiah Fuller, O. T. Clagett.** *Ibid.*, vol. 167, Dec. 1951, pp. 838–39.

12. Comparison of Aortic Pressure Pulses Recorded by Strain-gauge and Capacitance Manometer-Catheter Systems in Man, **E. H. Wood, I. R. Leusen, H. R. Warner.** *Ibid.*, vol. 171, Dec. 1952, p. 780.

Vol. 7
Pt. 1
1956

Germanium Rectifiers for Industrial Applications

L. W. BURTON
MEMBER AIEE

MANY processes and applications in industry require d-c power which has to be obtained from an a-c source. Metallic rectifiers have served well in this field of conversion, and the operating characteristics of magnesium copper sulphide, copper oxide, and selenium are well known. A more recent development is the germanium rectifier, with different characteristics which required that new design techniques be developed. A review of these characteristics in terms of the older metallic rectifiers will be a help in understanding the design and application problems.

First will be outlined the physical construction of the germanium rectifier in comparison with the other types of metallic rectifier cells and assemblies. The selenium and copper oxide rectifiers use large rectifying areas, operating at low current densities, such as 0.3 ampere per square inch. This means the use of large cell areas or stack assemblies to provide for large power supplies. However the large area provides a simple means of dissipating the heat where it is generated. It also provides a large heat sink or thermal capacity to take care of sudden current surges or short-time overloads. Likewise, voltage surges are usually taken care of easily as long as they are not maintained over a long period. So, in selenium and copper oxide there is a rectifying device of large-area low-current density, with well-established operating characteristics and, as a consequence of the experience gained over years of successful application and operation, it is usually considered quite simple to apply and operate.

By contrast, the germanium rectifier is exceedingly small and it operates at several hundred amperes per square inch of current density. However, it is much more efficient, which results in less heat to dissipate in total but still a large amount on a unit area basis.

The germanium crystal itself is a small wafer, 0.020 to 0.030 inch thick and up to 1 inch in diameter, although most of the rectifiers for power applications use a crystal about 1/2 inch in diameter. This crystal, in turn, becomes part of a sandwich, soldered between two metal plates which become collectors. The selection of these metals is important in the electrical and thermal characteristics obtained from the cell.

In the beginning, the material selected was ferni since its expansion characteristics were somewhat similar to that of the germanium crystal. This was then soldered to a suitable cooling surface, usually of copper. The ferni, however, was a poor heat conductor in the thermal circuit. A technique has now been developed for soldering the germanium crystal directly to the copper and thus greatly improving the heat transfer characteristic of the assembly.

Various methods of cooling are used in which the cell is mounted on a suitable heat transfer device depending on the particular method to be employed. Direct forced air cooling is used, usually, for lower power applications, with liquid cooling for larger power applications. Two types of liquid cooling are generally employed. One uses a cooling agent which is a dielectric, and this is pumped or circulated directly over the cell and fin assembly. It is then passed through a liquid-to-water heat exchanger. A more recent development is direct water cooling in which the cooling fins are located directly in the water-circulating system, and the cell is mounted outside. This has a high rate of cooling and greatly simplifies servicing since each cell assembly can be removed separately.

Another consideration in cell design is that of protecting the rectifier junction from moisture and other contaminants. This was done in the earlier designs by immersing the entire cell assembly in a neutral dielectric in which a drying agent is included. A more recent procedure is to use a complete, hermetically sealed, vacuum-tight assembly with ceramic insulation, and which is welded closed. This ensures complete isolation and no possibility of injury to the cell surface. On some of the early air-cooled units the cells were protected by potting them in epon plastics. This was satisfactory, but over a long operating life, which germanium can be expected to provide, the hermetic type of seal assures that there will be no change or damage to the cell from outside contaminants.

Some of the advantages and disadvantages of the germanium rectifier are listed in the following. The means taken to overcome the disadvantages will be discussed. The advantages are:

1. High operating efficiency.
2. Very good regulation.
3. Small size.
4. Low forward resistance.
5. High reverse resistance.

The disadvantages are

1. Low thermal capacity.
2. Temperature limitations.
3. Susceptibility to current surges.
4. Susceptibility to voltage surges.

One of the outstanding advantages of the germanium rectifier is its high operating efficiency of 98% for the cells, and an over-all equipment efficiency of 94% at 95% at an operating voltage of 65 to 75 volts direct current from a 3-phase power supply. Exceptionally good regulation is obtained as a result of the low forward drop in the cell. This, however, presents difficulties in applying these rectifiers where high current outputs are to be used and where a number of cells are to be operated in parallel. It has been done successfully with as many as 24 cells operating in parallel from a single transformer secondary in a 3-phase bridge circuit. This required close forward grading and matching of the cells as well as good design practices in making bus connections and bus layouts. A small resistance value could also have been added in series with each cell, but this would have lowered the rectifier efficiency slightly. In selecting cells to operate in parallel, it is necessary to grade them at their operating current and also to be certain of similar characteristics at overload currents.

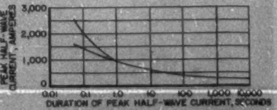


Fig. 1. Germanium cell overload current characteristics

Paper 56-78, recommended by the AIEE Metallic Rectifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 24, 1955; made available for printing December 14, 1955.

L. W. BURTON is with the General Electric Company, West Lynn, Mass.

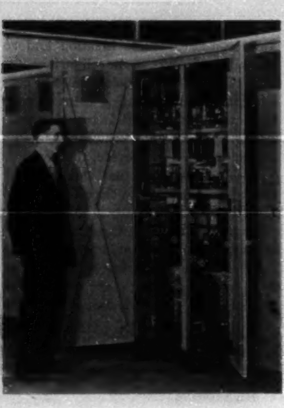


Fig. 2. Germanium rectifiers for anodizing: seven cubicles, 24 volts, 42,000 amperes

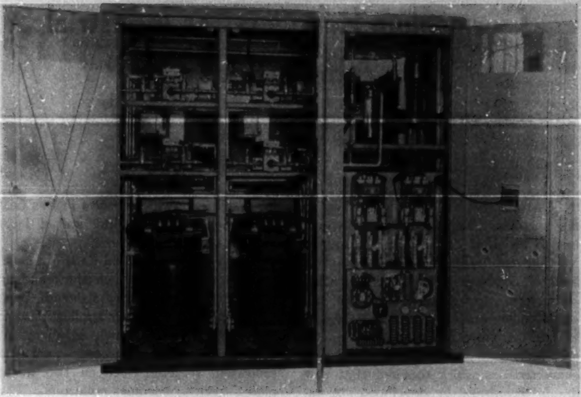


Fig. 4. Germanium rectifier single cubicle rated 60 volts, 2,000 amperes

The small size of the germanium rectifier is not only an advantage but also carries the disadvantage in terms of small thermal capacity. This shows up on overloads caused by internal faults and external short circuits.

The fault or overload current capabilities of the cell and cell assembly can be accurately calculated, and the calculations accurately verified, by precision laboratory tests. These values will vary with cell size, geometry, and materials. When the assembly is designed to give maximum heat transfer, the cell characteristics of current versus time will be similar to those shown as the top curve in Fig. 1. The cell illustrated here would have a normal continuous rating of approximately a 250-ampere peak, but could carry up to 900 amperes for 1 second, or 2,500 amperes for 0.04 second. The lower curve of the figure represents another cell assembly with less efficient heat transfer.

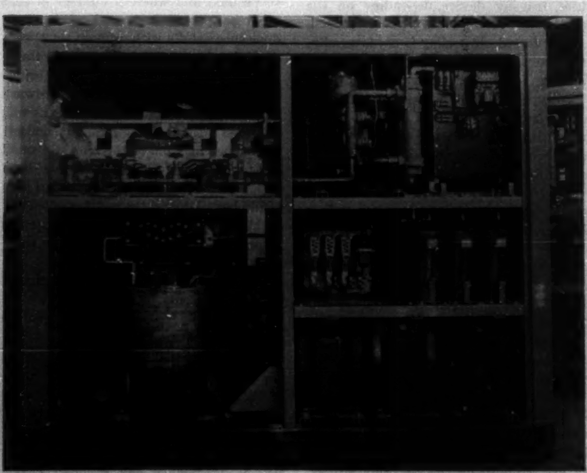


Fig. 3. Germanium rectifier single cubicle as used in Fig. 2

The limited thermal capacity of the germanium cell then brings up the problem of proper protection of the cell. It is obvious that standard fuses or circuit breakers operate too slowly to protect the cells. Also, if a circuit breaker is used with a group of cells operating in parallel, and one cell should fail (short-circuit out), the other branches would then start feeding short-circuit current to the failed cell and carry overloads that would cause other cells to fail before standard fuses or a circuit breaker could clear the fault. This could lead to the failure of as many as half the cells in the circuit.

The logical approach is then to develop a protective system for each cell, which will operate to limit fault current to a value less than that indicated as the time-current characteristics of the cell assembly (see Fig. 1). A fast-acting current-limiting fuse has been developed for this protection. It is usually placed in series with each cell, although satisfactory operation and protection have been obtained with two cells operating in parallel with one fuse. The voltage drop through the fuse is low, but it does help somewhat in balancing division of current when operating a number of cells in parallel. In circuits requiring lower current output where cells are not operated in parallel, such as a (6-1-1-*B*) 3-phase bridge, satisfactory protection can be obtained by placing the fuses in the a-c input circuit to the cells. In this case, if two or more rectifiers are operated in parallel, it is desirable to place a fast-acting fuse in the d-c output to interrupt d-c feed into an internal fault in one rectifier.

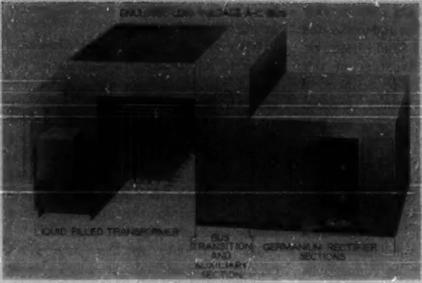

LOW VOLTAGE A-C BUS
LIQUID FILLED TRANSFORMER
BUS TRANSITION AND AUXILIARY SECTION
GERMANIUM RECTIFIER SECTIONS

Fig. 5. Germanium rectifiers for gas production; 65 volts, 42,000 amperes

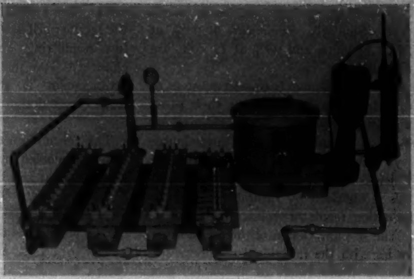


Fig. 6. Germanium rectifier cooling system

The use of a large number of fuses presents the problem of proper monitoring or signalling in case of the operation of one or more fuses. There are a number of systems which can be used, such as scanning for a voltage across each fuse at predetermined intervals, for example, once a minute. This signal can then indicate that a fuse has failed, and it can also sound an alarm.

Another characteristic of germanium which must be considered in applications is failure of the cell when transient voltages are encountered above the peak inverse voltage of the cell. The cell is destroyed when this occurs under normal operating conditions. Unhappily, it is almost impossible to apply germanium rectifiers to a system without having to contend with voltage transients. In any circuit where inductance, resistance, or stored energy in some form exists, the chances of a transient voltage are excellent. More specifically, they exist when switching inductive circuits or in energizing a system composed of inductance and resistance. The detection of a very short time-voltage transient is difficult. An electronic surge recorder or an oscilloscope with a high accelerating potential, and patience, are usually needed.

Two methods of handling this condition are used. One is to assign a voltage rating to the cell having sufficient safety factor to take care of normal surge conditions. Under normal conditions these surge voltages can run to three times the working peak voltage. The other method is to absorb these transients, since the energy content is usually low. A capacitor can normally be used to limit the magnitude of these short-time transients. If a surge counter is used, the magnitude of the capacitor may be determined by "trail and success." The capacitor size is increased until no surges above the safe value are detected. When an oscilloscope is used, the energy and magnitude of the transient can be determined, and the size of the capacitor calculated.

Another condition of operation that must be considered is the division of blocking voltage when cells are operated in series for higher voltage applications. One of the best ways of doing this is to add a small resistor across each cell. Although this decreases the efficiency slightly, it is the more acceptable method. The value of the resistance will depend upon the degree of voltage balance required and the characteristics of the cells.

The ratings of the germanium cell and the cooling system are designed to limit the cell temperature to a safe operating value. As the temperature of the cell increases, the peak inverse voltage of the cell decreases, so there is danger of cell failure if it is operated above the rated temperature rise.

From the foregoing discussion may be listed a number of conditions to be considered in designing and applying germanium rectifiers as follows:

1. The amount of overload current expected under short-circuit conditions.
2. The time the cell must withstand the overload, or the time it takes the protective device to operate.
3. The steady-state forward current.
4. The applied inverse voltage the cell must withstand.
5. The transient voltage the cell must withstand.
6. The maximum ambient temperature.

All these conditions must be considered in the designing and the use of germanium rectifiers; for example, it may be found that the surge current is more limiting than the continuous current rating, and it is necessary to operate the cell at less than its normal continuous rating. Also, there may be conditions of operation such as high ambient temperature or prolonged overloads that will increase the cell temperature above the normal operating limit.

How some of these germanium rectifiers have been applied will now be outlined. The majority of the applications now installed are for electrochemical processes requiring comparatively low voltage, high current, and high efficiency. These include applications such as the production of hydrogen, chlorine, and other gases, usually operating at 65 volts direct current. The electroplating and anodizing applications use from 12 to 48 volts direct current at high current. Power supplied to an arc reduction furnace usually required 80 volts d-c open-circuit with full-load operation at 40 to 60 volts. The tin electroplating of steel uses 15 volts, while electrocleaning uses 30 volts. All of these are desirable applications of germanium since the loads are fairly uniform with little chance of load short circuit. Efficiency is high with an over-all including all losses of 94% at

Table I. Typical Installations of Germanium Rectifiers

Application	Volts	Amperes
Electroplating	12	3,000
	12	45,000
	15	160,000
Electrolytic cleaning	30	10,000
Gas production	65	12,000
	65	16,000
	65	30,000
	65	42,000
Anodizing	24	10,000
	24	20,000
	24	40,000
	50	3,000
Arc furnace	40	4,000
	40	8,000
Total output all installations	16,000 kw	

$37^1/_2$ or 65 volts, depending on the circuit used. Similarly, efficiencies of 92% at 24 or 48 volts, and 90% at 12 volts are obtained.

The circuits have been selected so as to operate the cells at as near their rated voltage as possible. These circuits, all 3-phase, include the bridge and double wye with interphase. On applications in which control of the output is required, this is obtained by saturable reactor or induction regulator voltage control.

The size and design of the transformer varies with the individual requirements. Usually transformers of up to 300 kva have been designed and installed as part of the over-all cubicle. In some cases the cubicle includes the germanium stack assemblies, transformer, control reactor, a-c switchgear, fuse-monitoring system, and heat-exchanger type of cooling system.

For larger power requirements, these cubicles have been then connected in parallel. One such installation uses seven such cubicles to deliver 24 volts at 40,000 amperes for anodizing; see Figs. 2, 3, and 4.

Another procedure for a large block of power for a hydrogen cell line makes use of a large transformer similar to an arc furnace transformer with multiple secondaries. The rectifier banks contain only the rectifier cell assemblies and associated fuse-monitoring circuit and heat exchanger. The main advantage of this arrangement is the more economical transformer design that results; see Fig. 5.

The germanium cells in these applications are cooled by the direct, pumped liquid system and then a liquid-to-water heat exchanger. The rectifier cells are assembled in groups of 18 to 24, depending on the current required. They are then placed in a tank and completely sealed except for inlet and outlet connections for the coolant. Lead through connectors for both the alternating current and direct current are provided in the covers. The cooling system includes a reservoir and centrifugal pump which circulates the coolant at a comparatively high velocity through the trays and the heat exchanger. A drying agent is used in the coolant system together with proper temperature warning, and cutoff thermostats and flow indicator. Considerable care must be used in selecting the dielectric materials and gaskets because of the type of coolant used, since there is a tendency for the coolant to soften or attack many of the common type of insulating and gasketing material. A system layout is illustrated in Fig. 6.

With all the thought and discussion given to liquid-cooled systems, the air-cooled installations should not be forgotten. A number of 25-kw units delivering 65 volts are installed, operating in series parallel, for generating hydrogen, the reclamation of tin, and other applications. In these the air passes directly over the cooling fins. The main difficulty has been in keeping the fins clean so that proper cooling can be obtained.

These air-cooled rectifiers have given excellent service and certainly warrant consideration where water is not available for cooling, or for smaller power requirements. New fin designs now minimize the danger of clogged air passages.

A list of some of the applications now installed, or being installed, Table I, will indicate the extent to which germanium rectifiers supply d-c power to industrial installations. Their use and application are new, but no fundamental difficulties have been encountered. It seems certain that germanium rectifiers are desirable for industrial applications where power requirements are large, the output voltage comparatively low, and high operating efficiencies are desirable.

Principles of Noise Reduction in Communication Channels

L. S. SCHWARTZ
MEMBER AIEE

THE development of communication theory has led to the introduction of concepts such as integration, correlation, matched filtering, comb filtering, probability of detection, etc. Much confusion centers about the significance of these concepts because of functional equivalence and because of inadequately defined purposes in using them. Thus, it may be overlooked that integration and correlation are equivalent operations, or that matched filtering and comb filtering are techniques of integration. The purpose in using these concepts can be understood properly only within frameworks of suitable criteria of performance. For example, it is frequently assumed that the end object of a filtering technique is to improve the signal-to-noise ratio. Actually, improvement in signal-to-noise ratio is not an end in itself, but a means to an end. This end is to increase the information rate over a communication channel by reducing the probability of error.

Because of frequent confusion about the use of terms and the nature of the objectives and because of the importance of their proper understanding to rational system design, it seems desirable to discuss some basic objectives in the communications problem and to see how filtering concepts such as integration, correlation, matched filtering, etc., fall into perspective when considered within the framework of these objectives.

Communications Objectives

In communications, the transfer of a steady flow of information or a fixed amount of information from one point to another is of main concern. Communication services involving the use of teletype afford an example of a situation in which it is desired to maintain a steady flow of information. Radar is an example of a situation in which a relatively fixed amount of information is transferred, i.e., the position and/or velocity of a target at any instant.

For transfer of either a steady flow or fixed amount of information, maximum reliability is desired. In communication terms, transmission with maximum reliability is equivalent to transmission with minimum probability of error. The probability of error with which this paper is concerned is that resulting from the presence of noise in radio or radar re-

---

Paper 56-74, recommended by the AIEE Communication Theory Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30-February 3, 1956. Manuscript submitted October 17, 1955; made available for printing December 16, 1955.

L. S. Schwartz is with New York University, New York, N. Y.

This study was supported by the Evans Signal Laboratory, U. S. Army Signal Corps, under Contract *DA-36-039-sc-63133*.

ceivers. When the concept of minimizing the probability of error is applied to digital communication systems or to search radar, it is convenient to minimize the probability of error according to either one of two, but equivalent, approaches.[1] The first approach is the criterion approach and consists in guessing whether the signal is present or not according to some criterion. In the course of guessing, the operator may make one of two kinds of errors. He may report the signal present when, in fact, it is not (called a false alarm), or he may report it as not present when, in fact, it is (called a miss). The alternative to the criterion approach is to ask: What is the probability that a signal is present, if a noise-corrupted wave form in the output of the receiver is assumed? The first approach involves the concept of likelihood directly, and the second the concept of likelihood indirectly, via inverse probability. By "likelihood" is meant the probability that the receiver will respond to signal plus noise rather than to noise alone. The optimum receiver is the one which selects the value of input signal which maximizes the likelihood.[2]

Filtering Principles in Error Minimization

In both the criterion and inverse probability methods the receiver is the operator's analyzing system, enabling him to make inferences about the input from observations on the output. The operator can use the receiver output to advantage if the receiver outputs are different when a signal is and is not present, and if he can recognize this difference. This means that he must have information about the signal, about the noise, and about his receiver. The information about the signal and the noise is usually statistical because of the random nature of noise.

In the criterion method, the problem of detection is as follows.[3] There are available N-voltage samples which are known to have come either from a distribution of signal-plus-noise voltages or from a distribution of noise voltages alone. If the signal-and-noise properties are known, the probability density function of signal and noise and that of noise alone can be specified. From knowledge of the density functions, it is possible to compute the probability that all the voltage samples came from the distribution of signal plus noise and the probability that they all came from the distribution of noise alone. The ratio of the first probability to the second is the best measure or likelihood that all samples came from the signal-plus-noise distribution. Any monotonic function of this ratio gives an equally good significance test. A constant, which the ratio must exceed to say that the samples came from the signal-plus-noise distribution, is then picked arbitrarily. This constant determines the false alarm time. If the logarithm of the ratio of the probabilities is taken, it is seen that, in the very important case of Gaussian noise at the input to the envelope detector, the measure or likelihood function indicates the following steps: to take the logarithm of the incomplete Bessel function of order zero of the voltage sample, to add these logarithms for each sample, and to require the sum to exceed a certain value (the logarithm of the constant exceeded by the ratio of probabilities), called the threshold. In other words the optimum receiver has an envelope detector with the law

$$y = \log I_0(av) \tag{1}$$

where

I_0 = the modified Bessel function of zero order
v = the amplitude of the noise voltage envelope
a = the signal-to-noise power ratio, prior to detection

(It is noted that for small values of the argument the logarithm of I_0 can be replaced by the square of its argument. For large values of the argument, the logarithm of I_0 can be replaced by the argument to the first power. Thus, the log I_0 envelope detector functions as a square law device for weak signals and as a linear device for strong signals.) Moreover, this detector is followed by an integrator which sums successive voltage samples at the output of the detector. The sum is then applied to a threshold device which, in practice, is some form of trigger circuit that operates on an all-or-none principle.

If the receiver is designed according to the foregoing criterion, it will perform in such a manner as to maximize the probability of detection for a prescribed maximum acceptable probability of false alarm given by the threshold setting. It will thus be a receiver for minimizing the probability of error.

The procedure followed in the inverse probability method achieves the same results as the criterion method. Its characteristics are now briefly described. The former method developed by Woodward[2] proceeds from the point of view of "inverse" probability since the problem is to discover the cause x which has produced a given effect y, whereas "direct" probabilities describe the effects produced by a given cause. The product law for probabilities is

$$p(x)p_x(y) = p(y)p_y(x) \tag{2}$$

where

$p(x)$ = the probability that the signal x is transmitted
$p_x(y)$ = the probability that if x is transmitted y is received
$p(y)$ = the probability that y is received
$p_y(x)$ = the probability that if y is received, x was transmitted

$p_y(x)$ is the inverse probability, and may be written as

$$p_y(x) = kp(x)p_x(y) \tag{3}$$

where $k = 1/p(y)$, a constant, since, in the reception problem, it is assumed that y is given. The quantity $p_x(y)$ is known as the likelihood function and may be replaced by a quantity $L(x)$, so that

$$p_y(x) = kp(x)L(x) \tag{4}$$

To illustrate the significance of these concepts to communications, the example of a given received signal combined with additive Gaussian noise is considered. The reception problem is to operate on the received signal-plus-noise function in such a way as to maximize the probability that the additive mixture of signal and noise originated with some particular message at the sending end. The received voltage is

$$y = u_x + n \tag{5}$$

where x denotes the message and u_x the signal voltage representing it. For a Gaussian distribution of noise,

$$L(x) = \exp[-(y-u_x)^2/2N] \tag{6}$$

where N is the mean-squared value of n, and

$$p_y(x) = kp(x)\exp\left(-\frac{1}{N_0}\int_0^T (y-u_x)^2 dt\right) \tag{7}$$

for the received signal over a time T in a receiver of bandwidth W. The samples are taken $1/2W$ second apart, and N_0 is N/W, the mean noise power per unit bandwidth. Woodward[2] has shown that maximizing $p_y(x)$ is equivalent to maximizing the likelihood function $L(x)$.

It is observed that in equation 7, $L(x)$ and, therefore, $p_y(x)$ is maximized if the integral

$$\int_0^T (y-u_x)^2 dt \tag{8}$$

is equated to zero. If the signal is constant for the duration T, the operation performed in equating the integral to zero is one of averaging the noisy wave form. If u_x is a function of time and varies over T, the likelihood function will be maximized if the function

$$q(x)=\int_0^T yu_x dt \tag{9}$$

is maximized. It may be seen from the discussion on correlation which follows in a later section that the function $q(x)$ is a measure of the crosscorrelation between y and all the possible wave forms u_x. Thus, the most probable message state is the one which yields the largest crosscorrelation. It follows that in the case of Gaussian noise the optimum receiver is a crosscorrelation device.

At first sight it may appear that the results of the criterion and inverse probability approaches are in disagreement, because it is found from the application of the former that the ideal receiver is a log I_0 detector followed by an integrator, whereas it was just found from the application of the latter that it is a crosscorrelator. Actually, the two results are not in disagreement, since the methods were applied to two different situations. In the application of the criterion approach, an envelope detector was considered but, in the application of the inverse probability approach, nothing was said about the kind of detector that was used. Now, however, the results of applying the latter approach to a case involving an envelope detector will be studied. In the foregoing attention was focused on a quantity $q(x)$ which is a measure of the crosscorrelation between y and u_x. If x is a function of time over the interval of interest $q(x)$ may be replaced by $q(\tau)$; note that u_x and $n(\tau)$, the noise voltages, are radio-frequency functions of time. It follows that $q(\tau)$, itself must also be a radio-frequency function of time, with an envelope varying comparatively slowly. By a suitable choice of ω

$$q(\tau)=Q(\tau)\cos(\omega\tau+\phi) \tag{10}$$

where ω is a high frequency with $Q(\tau)$ and ϕ varies comparatively slowly with τ. These statements merely say that $q(\tau)$ possesses fine structure and, since the inverse probability $p_y(\tau)$ is a function of $q(\tau)$, it also will exhibit fine structure which needs to be removed. It can be shown[8] that the correct procedure is to smooth $p_y(\tau)$ in such a way that the area under $p_y(\tau)$ for each interval of a radio-frequency cycle is preserved. Thus, for the smoothed distribution

$$p_y(\tau)=kp(\tau)\int_{T-\pi/\omega}^{T+\pi/\omega}\exp[Q(\tau)\cos(\omega\tau+\phi)]d\tau \tag{11}$$

The integral is $I_0(Q)$ where

I_0 = the modified Bessel function of zero order

Q = the positive envelope of q

Hence

$$p_y(\tau)=kp(\tau)I_0(Q) \tag{12}$$

The process is detection with an I_0 characteristic. If observations are taken on successive voltage samples

$$p_y(\tau)=kp(\tau)\prod_n I_0(Q_n) \tag{13}$$

This may be written alternatively as

$$p_y(\tau)=kp(\tau)\exp\left[\sum_n \log I_0(Q_n)\right] \tag{14}$$

This result shows that the most probable message state is the one which yields the largest value of the sum of sample voltages in the output of a log I_0 detector. This is the result also yielded by the criterion approach.

The implications of the foregoing results are important. It is seen that if envelope detection is specified, the optimum receiver is a log I_0 detector followed by an integrator. If not, the optimum receiver is a crosscorrelation device. In view of the fact that envelope detection is used in conventional receivers, the question arises as to whether there are times when it is preferable not to use it. The answer is yes. In envelope detection, when the signal is smaller than the noise, the noise suppresses the signal because of the nonlinearity of the process, so that the signal-to-noise ratio at the output of the envelope detector is smaller than at the input. If it is required to operate under high relative noise conditions, it is advisable, therefore, to avoid envelope detection. The alternative to envelope detection is synchronous detection which is described in some detail later in this paper. Synchronous detection will be seen to be a method of removing signal fine structure by crosscorrelation. Since crosscorrelation is a linear process, suppression of weak signals by noise does not take place.

When the question of applying crosscorrelation is considered, it will be seen that alternative ways exist for achieving it: by linear integration, by "comb" filtering, and by correlation itself. The term "matched filtering" as used here applies to integration over a single signal duration. In pulse systems, such as radar and digital communication systems, integration is over a single pulse. A matched filter may be used in conjunction with an integrator, following an envelope detector, to supplement the improvement obtained by adding successive signals.

Improving Signal-to-Noise Ratio as a Means of Minimizing the Probability of Error

Implicit in the operation of optimum receivers is the existence of a threshold device at the output of the receiver. A threshold circuit is a necessity because a receiver is, in the last analysis, a decision device. It must decide whether a signal is present or not, and it must make this decision with minimum probability of error. The decision is a yes-no operation which is implemented as to circuits by means of the threshold device. The problem of decision then resolves itself into a question of whether the output of the receiver exceeds the threshold or not. If noise alone is present, the threshold should not be exceeded, and if signal plus noise is present, the threshold should be exceeded. Since noise alone can have any amplitude with finite probability, it is evident that noise voltages will occasionally exceed the threshold and cause triggering of system output circuits, no matter what the threshold. This event is called a false alarm, and it is desired to have its probability small, a condition which is achieved by raising the threshold sufficiently. The threshold cannot be raised indefinitely, however, because then the probability of misses would become unacceptably large. Accordingly, a compromise must be reached which represents an acceptable balance between the false alarm rate and the miss rate. On the other hand, if the signal-to-noise ratio at the input of the threshold were increased, the threshold could be raised for a given miss probability, or lowered for a given false alarm probability depending on whether the signal or the noise had been increased relative to the other. In either case, the result is a reduced probability of error, the object of increasing the signal-to-noise ratio. By application of either the criterion or inverse probability approaches, it can be seen that integration or correlation, which are techniques of linear filtering, are optimum means for minimizing probability of error when the noise is Gaussian. Since the probability of error is a function (exponential) of signal-to-noise ratio, it follows that linear filtering techniques are also optimum methods of improving signal-to-noise ratio when the noise is Gaussian. For this reason, it is desira-

Vol.
Pt. 1
1956

ble to compare briefly the various techniques commonly used to improve signal-to-noise ratio and to show the relationship they bear to each other. These techniques are: matched filtering, comb filtering, linear integration, and correlation. Each of these will now be discussed.

MATCHED FILTERING

The matched filter, first considered by North[4] and then later and independently by Middleton and Van Vleck,[5] is a filter designed to achieve maximization of the signal-to-noise ratio. The figure of merit ρ, on which the filter design is based, is defined as the ratio.

$$\rho=\frac{|v(t_0)|^2}{\sigma^2} \qquad (15)$$

where

$v(t_0)$ = the value of the output signal at some specified instant $t=t_0$

σ^2 = the mean-square value of the output noise

Application of Schwartz's inequality to this ratio shows that the transfer characteristic of the filter which maximizes ρ is the complex conjugate of the spectrum function of the input signal. Such a filter has the property that its impulse response is the image of the input signal. The output of the matched filter for a rectangular pulse input is a triangular pulse the base width of which is twice the width of the input pulse. It is observed that the output of a correlation device for a rectangular pulse input is also a triangular pulse of base width equal to twice the input pulse width. Also, time-aperturing (which is convolution, an integration process) a rectangular pulse with an impulse response function which is the image of the input pulse results in a triangular output. Thus, the matched filter can be looked upon either as a correlation or an integration device. This observation will be supported by a mathematical statement in a later section dealing with correlation.

In determining the improvement accomplished by a matched filter, its performance may be compared with that of an resistance-capacitance (R/C) filter the bandwidth of which is appropriate to the signal being used. For a rectangular pulse in Gaussian noise, the maximum value of ρ is

$$\rho_{MF}=\frac{V^2\delta}{N_0} \qquad (16)$$

where

V = the amplitude of the rectangular pulse

δ = the pulse width

N_0 = the value of the noise power spectrum

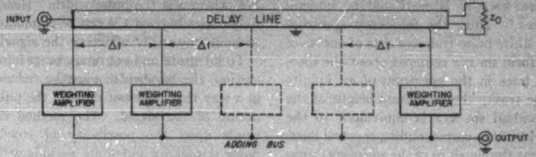


Fig. 1. Diagram of a matched filter

For an R/C filter, the output signal-to-noise power ratio is

$$\rho_{RC}=\frac{V^2\delta(1-\epsilon^{-2\pi\alpha})^2}{\alpha\pi N_0} \qquad (17)$$

where α is a factor which determines the relation that the frequency cutoff point of the R/C filter bears to the inverse of the pulse width. The other quantities have been defined previously. The improvement factor in signal-to-noise power ratio for a matched filter over an R/C filter is defined as

$$R=\frac{\rho_{MF}}{\rho_{RC}} \qquad (18)$$

$$R=\frac{\alpha\pi}{(1-\epsilon^{-2\pi\alpha})^2}. \qquad (19)$$

It is observed that the relative improvement of the matched filter over the R/C filter increases with α. This means that if the performance of the matched filter is compared with that of an R/C filter the cutoff frequency of which is appreciably larger than $1/\delta$, where δ is the pulse width for which the matched filter is designed, the performance will be better than if the comparison is made with an R/C filter of smaller relative cutoff frequency. The physical basis of the improvement is that the matched filter has a frequency transfer function which resembles the spectrum of the input signal so that, unlike a nonoptimum filter, it admits only that noise which is coincident with the spectral components of the signal or which lies between them, but not the noise which lies outside the signal spectrum.

It might be considered that a matched filter would be designed in the frequency domain since theory yields the frequency transfer function explicitly. The equivalent filter design can be achieved much more easily, however, in the time domain from knowing that the impulse response is the convolution of itself with the periodic unit impulse as scanning function. Thus the impulse response is given by

$$W(t)=\int_{-\infty}^{\infty}W(\tau)u(t-\tau)d\tau \qquad (20)$$

and

$$W(t)\backsimeq\sum_{n=0}^{N}W(n\Delta t)p(t-n\Delta t) \qquad (21)$$

where n is the total number of discrete time intervals of width Δt. In equation 21, n is summed from zero because the realizable impulse response is zero for $t<0$, and $p(t)$ is a pulse of duration Δt, approximating the impulse. Since the impulse response of the matched filter is the image of the input signal, the form of $W(t)$ is specified. The basic components of the matched filter include a tapped delay line, weighting amplifiers, and an adding bus. A block diagram of a matched filter is shown in Fig. 1.

COMB FILTERING

The comb filter, unlike the matched filter, integrates over more than one pulse. Comb-filter operation is tantamount to insertion of band-pass filters of infinitesimal width at the spectral lines of a repetitive signal. Thus, the essential frequency components of the signal are admitted by the filter whereas the noise which fills the region between spectral lines is rejected. If the band-pass filters are infinitely narrow, the filter, in effect, operates on an infinite pulse train, and the filtering is perfect. In practical applications, perfect operations is not possible, and comb filtering must be limited to a finite number of pulses. In view of this, it is necessary to consider the shape of the spectrum of a finite group of pulses as distinct from the spectrum of a succession of repetitive pulses of infinite extent.

The frequency spectrum of a signal consisting of a train of N rectangular pulses of constant width and amplitude and period T is

$$S(\omega)=\sum_{k=-\infty}^{\infty}F\left(\frac{2\pi k}{T}\right)\times\frac{\sin NT\left(\omega-\frac{2k}{T}\right)}{\frac{NT}{2}\left(\omega-\frac{2\pi k}{T}\right)} \qquad (22)$$

where $k=n/N$ is restricted to integral values.[6] Thus, the frequency spectrum of a finite pulse train consists of curves of the form $\sin x/x$ centered about the spectral lines in the spectrum of an infinite pulse train; the maximum height of the individual $\sin x/x$ is equivalent to the height of the corresponding spectral lines, so that the envelope of the entire spectrum is also of the form $\sin x/x$. As N, the number of pulses in the train, increases, the individual $\sin x/x$ curves about the spectral lines become increasingly spiked, and the spectrum approximates the line spectrum of an infinite train of pulses.

The improvement obtainable with a comb filter increases with the number of pulses on which the comb filter operates, because, as just observed, the spread of the $\sin x/x$ function about the spectral lines decreases as N increases, admitting less noise to the output of the filter. Since the transfer characteristic of an ideal (linear) integrator consists of a series of spikes at the spectral lines, the transfer characteristic of a comb filter approximates that of an ideal integrator as N increases. Because of the functional similarities between the comb filter and the linear integrator and because of the technical difficulties involved in achieving the necessary frequency stability for satisfactory operation of the comb filter, its use is not recommended.

LINEAR INTEGRATION

Integration and correlation have been widely described in the literature and have been employed to aid in the detection of periodic signals obscured by noise. The advantage of integration or correlation consists in the fact that signals reinforce upon successive additions in time, whereas noise both reinforces and cancels. It is evident that the best result is obtained when the noise has minimum correlation with itself, i.e., when it has the statistical structure of thermal noise.

Integration may be achieved by introducing the signal wave form into a number of storage elements with given delays and then summing the delayed signals. Ideally, an infinite number of pulses would be summed to obtain an infinite improvement in signal-to-noise ratio. A far from ideal example of a linear integrator is a low-pass filter which integrates by eliminating high-frequency components in the noise. Signal frequencies, however, may also be eliminated if the bandwidth of the filter is reduced beyond the point at which the signal spectrum has significant energy. For pulsed signals, as in radar, the signal spectrum has little energy at frequencies greater than $1/\delta$, where δ is the pulse width. Hence, frequencies beyond $1/\delta$ can be eliminated without materially affecting the signal.

To integrate and yet retain range information, the bandwidth must be reduced in a way that does not destroy the pulse nature of the signal. The method depends upon the possibility of storing radar signals for indefinite periods without distortion. Each repetition period is treated as a unit, and its signals are stored in such a way that they can be recovered in their correct time sequence. The combined signal, with pulses from fixed targets correctly superimposed, is given by

$$S(t)=\sum_{n=0}^{\infty} a_n s_n(t-nT) \quad 0<t<T \tag{23}$$

The change in signal-to-noise ratio may be calculated by considering a particular instant. If this instant lies within a signal pulse, the signal amplitudes will add to a sum proportional to $\sum_{0}^{\infty} a_n$. The resultant rms noise voltage on the other hand will be proportional to

$$\left(\sum_{0}^{\infty} a_n^2\right)^{1/2}$$

since the noise is uncorrelated between successive repetition periods. The improvement in signal-to-noise power ratio is therefore

$$R_I=\frac{\left(\sum_{0}^{\infty} a_n\right)^2}{\left(\sum_{0}^{\infty} a_n^2\right)} \tag{24}$$

It is clear that the signals are summed with a weighting function which, in the case of the linear integrator for radar pulses, has peaks in the repetition frequency and its harmonics. The frequency characteristic of the integrator is therefore of the comb-filter type, passing the Fourier components of the pulse train and attenuating noise at other frequencies.

CORRELATION

The subject of correlation is concerned with relations which exist between parts of the same signal, or between parts of one signal and parts of another. The former is defined by the autocorrelation function whose symbolic expression is

$$\psi_{11}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} f_1(t)f_1(t+\tau)dt \tag{25}$$

Physically, the autocorrelation function is simply the average product obtained by comparing the given signal with itself. If the signal is completely random, different parts of the same signal will be independent of each other and, consequently, the value of the autocorrelation will be zero except for $\tau=0$, meaning that the signal is correlated with itself only. An example of a completely uncorrelated signal is thermal noise, the autocorrelation function of which is a spike at $\tau=0$. For more correlated time functions, the autocorrelation function decays to zero for increasing τ.

The crosscorrelation function differs from the autocorrelation function in that it represents the average product relationship which exists between two different time functions and is thus a measure of the amount of correlation which exists between the functions at different intervals of time. Crosscorrelation is applied when a priori information concerning the frequency of the signal is known. If there is no a priori information, no advantage is obtained over autocorrelation.

It is instructive to consider the relationship between correlation and integration. If a signal $S_i(t)$ and noise $N(t)$ are applied to a filter with an impulse response $W(t)$, the filter output $S_0(t)$ is given by the convolution integral

$$\begin{aligned}S_0(t)&=\int_0^{\infty}[S_i(\tau)+N(\tau)]W(t-\tau)d\tau\\&=\int_0^{\infty}S_i(\tau)W(t-\tau)d\tau+\int_0^{\infty}N(\tau)W(t-\tau)d\tau\end{aligned}$$

If $W(t)$ is equal to $S_i(-t)$, the condition for a matched filter, the filter output is

$$S_0(t)=\int_0^{\infty}S_i(\tau)S_i(\tau+t)d\tau+\int_0^{\infty}N(\tau)S_i(\tau+t)d\tau$$

The similarity between this latter expression and the crosscorrelation function

$$\begin{aligned}\phi_{12}(\tau)&=\lim_{T\to\infty}\frac{1}{T}\left[\int_0^{\infty}S(t)S(t+\tau)dt+\int_0^{\infty}N(t)S(t+\tau)dt\right.\\&=\lim_{T\to\infty}\frac{1}{T}\psi(\tau)\end{aligned} \tag{26}$$

obtained by crosscorrelating the stored signal $S(t)$ with the same input signal, is immediately apparent. With the exception of the averaging process, the two methods of detection are similar. The output from the filter $S_0(t)$ is a function of t, whereas $\psi(\tau)$ is a function of τ. Thus

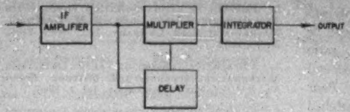


Fig. 2 (left). Block diagram of autocorrelation radar

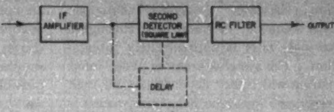


Fig. 3 (right). Block diagram of conventional radar

two types of filters may be used to obtain crosscorrelation in real time. It should be noted that the requirement that $W(t)$ be made equal to $S(-t)$ is equivalent to the requirement that the transfer characteristic of the filter be the conjugate of the Fourier transform of the signal $S_i(t)$. It will be recalled that this is precisely the condition for maximizing the peak signal-to-rms noise ratio.

If a sinusoidal function is crosscorrelated with a periodic signal possessing more than one frequency, one of which is a frequency equal to that of the sinusoidal function, the output of the correlator is a sinusoidal function of the same frequency as the correlating sinusoid. This is the same result as would be obtained if the periodic signal were passed through an infinitely narrow filter centered on the frequency of the correlating sinusoid. Since this discussion is limited to theoretically ideal correlators for which the time of integration is infinite, the correlator can be regarded as the equivalent of an infinitely narrow filter in the frequency domain. It is noted that the latter filter would have an infinite build-up time. Although correlation has its frequency domain analogue, such that the time and frequency domain viewpoints are operationally equivalent, it may be easier to build a highly stable oscillator than a correspondingly narrow bandwidth filter. If a periodic signal with harmonic components is correlated with another periodic signal possessing an infinite set of harmonic components of equal amplitude, the output of the correlator will consist of discrete frequency components representing the original signal. Thus, the ideal correlator behaves like a comb filter.

In summary the four linear techniques discussed in the foregoing are equivalent for signals in Gaussian noise in that they all maximize signal-to-noise ratio by linear integration and therefore minimize the probability of error associated with decision making.

Applications of Linear Filtering (Correlation)

Because of the importance of linear filtering techniques and because all linear techniques may be considered a form of correlation, it is worth while to look at some well-known applications to see if their noise-reducing properties can be attributed to correlation.

SYNCHRONOUS DETECTION

Conventional envelope detection utilizes only amplitude information and discards phase information, thus according to information theory, throwing away the equivalent of 3 decibels in signal-to-noise ratio. Much more serious, however, is the fact that envelope detection is a nonlinear operation which suppresses the signal if the noise exceeds the signal in amplitude. Thus, if the signal is 10 decibels below the noise at the input to the second detector, it will be 20 decibles below the noise at the output.[7] Also, because of the alteration in the statistical properties of the noise in the second detector, averaging filters are not as effective at video frequencies as they are at intermediate frequencies. From a practical standpoint, this proves to be unfortunate because the technical problems in filtering at intermediate frequencies are much more difficult than at video frequencies. Therefore, it is highly desirable to find an alternative to envelope detection.

Synchronous detection seems to be a promising alternative. Its properties are as follows. Transmitter and receiver frequencies are coherent so that the system uses phase information. Hence, it provides a 3-decibel signal-to-noise improvement over simple envelope detection. Secondly, the detection process is linear so that no suppression of the signal by the noise occurs for signals smaller than noise; since filtering operations can be conducted with equal efficiency before and after detection, practical difficulties in filtering at intermediate frequencies can be avoided.

The operation of synchronous detection is as follows. The message function $f_m(t)$ is modulated into a cosine or sine carrier at the transmitter to give

$$f_s(t)=f_m(t)\cos\omega_0 t \tag{27}$$

At the receiver the incoming signal is multiplied by a cosine or sine reference voltage of the same frequency as the carrier, the resultant voltage being

$$\begin{aligned} f_s(t)\cos\omega_0 t &= f_m(t)\cos^2\omega_0 t \\ &= \frac{f_m(t)}{2}(1+\cos 2\omega_0 t) \end{aligned} \tag{28}$$

The signal is then passed through a low-pass filter which eliminates the carrier frequency term. The output is $f_m(t)/2$. Synchronous detection is thus seen to be a linear means for removing fine structure from a received signal. To the extent that it employs multiplication with a known frequency and averaging, it resembles crosscorrelation, but it is not complete crosscorrelation because a low-pass filter is used which permits all essential harmonic components to be passed. A true correlator would permit the carrier frequency term only to be passed and thus would function as an infinitely narrow band filter.

CORRELATION IN RADAR

Pulse-Amplitude-Modulated Radar

The pulse-amplitude-modulated radar with a square-law second detector bears a striking resemblence to the autocorrelation system shown in Fig. 2 since it is really the conventional system in Fig. 3 with a delay network added. If the input signal to the circuit of Fig. 2 is a rectangularly shaped pulse, the output is a triangularly shaped pulse. If the input to the second detector in Fig. 3 is

$$f_1(t)+f_1(t+\tau) \tag{29}$$

the detector output is

$$[f_1(t)]^2+2f_1(t)f_1(t+\tau)+[f_1(t+\tau)]^2 \tag{30}$$

The R/C filter is a low-pass filter which functions approximately as an integrating device. Hence, the output of the detector is integrated and this results in

$$2\phi_{11}(0)+2\phi_{11}(\tau), \tag{31}$$

two autocorrelation functions. Because $\phi_{11}(0)$ is constant for a stationary process, the first term can be eliminated by calibration. For rectangular input pulses, $\phi_{11}(\tau)$ is a triangular function of time and thus it is seen that, with a slight modification, the pulse radar receiver can be modified to perform as an autocorrelation detector.

Continuous-Wave and Frequency-Modulated Radar

Continuous-wave doppler and frequency-modulated radars are fundamentally correlation systems since they heterodyne, i.e., multiply the received signal and a locally generated time function in the mixer. Passage of the doppler beat frequency or frequencies through a low-pass filter is approximately integration. The systems preserve phase information which they depend upon for their reading. Thus, the detection process is approximately crosscorrelation.

Conclusions

The basic objective in reducing noise in communication channels is to minimize the probability of error. The probability of error can be minimized by maximizing the likelihood ratio which, in terms of a physical operation, means improving the signal to noise. It is necessary to consider the way in which the signal-to-noise ratio is improved because there is an optimum way. In the extremely important case of Gaussian noise it can be shown that the signal-to-noise ratio should be improved by one of several possible linear filtering techniques. Four well-known techniques are: 1. matched filtering; 2. comb filtering; 3. linear integration; and 4. correlation. It is shown that all four of these techniques have equivalent properties and examples of the use of linear filtering techniques, specifically from the point of view of correlation, are given.

References

1. The Theory of Signal Detectability I, II, **W. W. Peterson, T. G. Birdsall.** *Technical Report No. 13*, EDG, University of Michigan, Ann Arbor, Mich., June, Jul. 1953.

2. Probability and Information Theory with Applications to Radar (book), **P. M. Woodward.** McGraw-Hill Book Company, Inc., New York, N. Y., 1953.

3. A Statistical Theory of Target Detection, Mathematical Appendix, **J. E. Marcum.** *Report No. RM-273*, Rand Corporation, Jul. 1, 1948.

4. Analysis of the Factors which Determine Signal-to-Noise Discrimination in Radar, **D. O. North.** *Report PTR-6C*, Radio Corporation of America Laboratories, New York, N. Y., June 1943.

5. A Theoretical Comparison of the Visual, Aural and Meter Reception of Pulsed Signals in the Presence of Noise, **J. Van Vleck, D. Middleton.** *Journal of Applied Physics*, New York, N. Y., vol. 17, Nov. 1946, pp. 940–71.

6. Comb Filters for Pulsed Radar Use, **S. F. George, A. S. Zamanakos.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, Jul. 1951, p. 1159.

7. Appendices to Comparison of Modulation Methods, **S. F. George.** *National Convention Record, Ibid.*, pt. 8, Information Theory, Mar. 1953, pp. 21–25.

Application of Statistical Techniques to Electron Tubes for Use in a 4,000-Mile Transmission System

W. VAN HASTE
NONMEMBER AIEE

OVER the past several years the subject of reliability of electronic equipment has received considerable attention. Much has been written on the need for reliable systems, and the mathematical probabilities of component performance have been explored. While there have been demonstrations of achievement, it is felt that additional material of this nature will be of interest. One of the chief factors in establishing a reliable system is to assure compatibility of the operating environment and the capabilities of the active electronic device. This can best be brought about by close cooperation between the equipment and component designers.

The purposes of this paper are: 1. to give an example of the achievement of reliability in a 4,000-mile communication system and 2. to discuss some of the statistical techniques and co-operative measures used to accomplish the feat. The order of presentation relates the system objectives to tube performance requirements. The latter are in turn dependent upon the dimensions of the electrodes that comprise the tube. The control of tube parts, the resultant effect on tube characteristics and, finally, the extent to which reliable system performance has been achieved are discussed.

System Objectives

The *L-3* coaxial system was designed by the Bell Telephone Laboratories to meet telephone and television requirements in the growing demand for long-distance transmission facilities. One transmission circuit is capable of handling 1,860 telephone message channels, or one television channel and 600 telephone message channels. The transmission band is 300 kc to 8.35 megacycles and requires amplification approximately every 4 miles. In a 4,000-mile system there are thus about 1,200 amplifiers in tandem, where each amplifier employs five electron tubes as active devices.

To have a practical maintenance procedure, it is necessary to use plug-in amplifiers which, in turn imposes a requirement of complete interchangeability. To achieve this goal within the tight system limits, it is essential that the amplifiers follow a normal distribution pattern. The economical way to secure this objective is by imposing distribution requirements on the amplifier components. In the following discussion it should be remembered that, while only the electron tubes are discussed, similar considerations apply equally well to the coils, condensers, resistors, and other circuit components.

The achievement of the following broad general system objectives is dependent on the satisfactory performance of the electron tubes:

1. The resulting transmission line must be equalizable.
2. The signal should be transmitted with a sufficient degree of fidelity.
3. The system must be reliable.

Electron Tube Requirements

The relation of the basic system objectives to the electron tubes, shown in Table I, is as follows:

1. Equalization: Equalization is defined as the ability to maintain amplitude and phase of the transmitted signal at the receiving terminal within limits determined by the facility being transmitted (i.e., telephone, telegraph, telephoto, or television). This requires that the average values of transconductance of the tubes

---

Paper 56-86, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted April 20, 1955; made available for printing November 21, 1955.

W. Van Haste is with the Bell Telephone Laboratories, Inc., Allentown, Pa.

The development described in this paper was accomplished by co-operation of many people working within the following branches of the Bell System: American Telephone & Telegraph Company, Bell Telephone Laboratories, and the Western Electric Company. Special mention is made of the contribution to the preparation of this paper by B. J. Kinsburg of the Bell Telephone Laboratories.

Table I. Relation of General System Objectives to Tube Requirements

System Objectives	Tube Requirements
Line must be equalizable	transconductance must be controlled initially and with time
Fidelity of transmission must be maintained	harmonic distortion must be low
Reliability must be high	long thermionic life, minimum sudden failures

be maintained close to desired nominals, that individual values be randomly distributed around these nominals, and that the reduction of transconductance with time be gradual (less than 3 decibels (db) in 15,000 hours).

2. Fidelity: The fidelity of transmission requires at least a 40-db signal-to-noise ratio. This may be translated into the electron tube harmonic-distortion requirements. While statistical controls have been placed on these measurements, they will not be discussed in detail.

3. Reliability: The over-all reliability requirement for the 4,000-mile system is that not more than one major failure (loss of circuit for 4 hours or more) shall occur per year. Preliminary studies showed that such reliability with available components could only be achieved by providing automatic switching to a standby circuit. Assuming automatic switching, the over-all reliability objectives for electron tubes of the *L3* system are:

a. Sudden failures (glass cracks, open connections, interelectrode short circuits, etc.) shall not exceed one per 1,000 sockets per month.

b. Average thermionic life should be at least 15,000 hours.

To meet the demands of this 4,000-mile transmission system, three new electron tubes were developed by the Bell Telephone Laboratories for use in the amplifiers. These are manufactured by the Western Electric Company under code numbers *W.E. 435A*, *W.E. 436A*, and *W.E. 437A*. All of these are high-transconductance tubes, the first two being tetrodes and the last one a triode. Table II lists the important characteristics of the three tube codes under typical operating conditions.

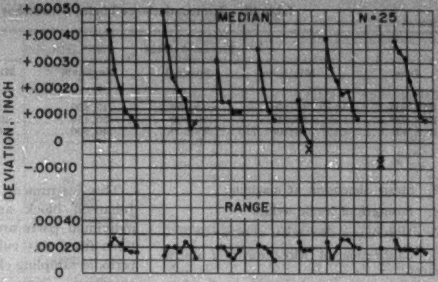


Fig. 1. Deviation of grid side rod diameter from desired nominal

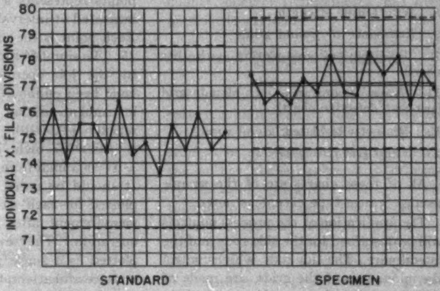


Fig. 2. Measurement of fine grid wire

Mention has been made of the need for a normal distribution of amplifiers to achieve interchangeability and also that the electron tubes used as the amplifying devices must be normally distributed in regard to some major characteristic. It is interesting to note that the 1,200 amplifiers in tandem actually work to the advantage of the manufacturers in that a normally distributed product is acceptable. In a very small system this might not always be true.

Electron Tube Manufacture

The maintenance of distributed performance of electron tubes is directly dependent on the piece parts from which the tubes are assembled. Process controls are maintained on the physical dimensions of the cathodes, grids, anodes, etc., and also on the effects of chemical processes such as cleaning, coating, heat-treating, etc. For the new tubes for the *L-3* system, virtually every part and process has statistical controls applied to it.

The control grid will be discussed in some detail since it is one of the most critical electrodes and, being unconventional, offers some novel features. A conventional grid is a self-supporting helix of lateral wire wound around two side rods. The wires are fixed in position by notching and swaging operations. To obtain the proper characteristics for the *W.E. 435A*, *W.E. 436A*, and *W.E. 437A* tubes, it is necessary to use grid lateral wire of approximately 0.0003 inch in diameter. A grid wound with such wire would not be self-supporting. The answer to the grid structure problem is a supporting frame made from two large-diameter side rods joined together by cross straps at the ends of the grid. The fine lateral wires are wound on this frame. The control charts being plotted in connection with grid manufacture are given in the following:

1. Diameter of side rods—as received.
2. Diameter of side rods—after tumbling.

Table II. Tube Characteristics

	W.E. 435A Tetrode	W.E. 436A Tetrode	W.E. 437A Triode
Transconductance, micromhos	15,500	30,000	46,000
Plate current, milliamperes	11.4	20.2	36
Screen current, milliamperes	2.9	8.2	
Capacitances (no shield), micromicrofarads			
C_{gl-p}	0.02	0.04	3.6
$C_{gl-k+gs}$	7.1	14.0	11.0
C_{p-k+gs}	1.02	1.6	0.74
Figure of merit $\left(\frac{g_m}{C_{in}+C_{out}}\right)$	1,500	1,640	3,810

Table III. Transconductance Limits, Micromhos

Tube	Process Average			Individuals	
	Minimum	Nominal	Maximum	Minimum	Maximum
W.E. 435A	15,200	15,500	15,800	12,500	18,500
W.E. 436A	29,550	30,000	30,450	25,500	34,500
W.E. 437A	45,400	46,000	46,500	40,000	51,000

3. Strap dimension at machine.
4. Strength of frame welds.
5. Minor axis dimension at straps.
6. Calibration of wire-measuring instrument.
7. Diameter of bare lateral wire.
8. Wire tensile strength.
9. Resonance of wound grid.
10. Efficiency of gold-plating bath (bar chart).
11. Visual appearance of plated grid (per-cent defective chart).
12. Diameter of plated lateral wire.
13. Final grid resonance.

All of these charts are of the sample average–range type except as noted.

The application of the process controls will be described for the side rods and the lateral wire. The side rods, which control the minor axis of the grids, are made of molybdenum and are required to be held to a tolerance of ±0.0001 inch on diameters in the order of 0.040 inch. The parts are purchased oversize and are literally tumbled down to the proper size in an abrasive solution. An analysis of variance demonstrated that the measuring method is of great importance in obtaining results of the proper significance. Use of barrel micrometers and dial-indicating gauges were found inadequate and side-rod diameter measurements are now made on a supermicrometer.

This instrument is calibrated with a Johansen block and measurements of individual parts are recorded as deviations from this value. Fig. 1 shows a periodic sampling chart for the tumbling operation. Each group of connected points represents a lot of tumbled rods measured at intervals during the process. Each point is the median of a sample of 25. To simplify the handling of numbers by the job operators, negative values have been eliminated by calibrating the instrument so that a nominal part will register +0.00010. The process average is +0.00007 inch, indicating a product −0.00003 inch from the desired nominal. It should be noted that the final point plotted for each set of rods represents a measurement after a chemical cleaning operation. The seventh lot of rods is plotted as a single point, illustrating an unsuccessful attempt to eliminate periodic measurements. Horizontal lines on this chart represent objective working limits and not process control limits.

Another intricate problem of measurement is encountered with the grid lateral wire which is nominally 0.00029-inch ±0.00001-inch-diameter tungsten (etched to the proper diameter). Complication of this measurement is introduced by requiring an accurate figure on the thickness of the gold that is plated on the wire. The precision demanded for this work taxes the available instruments as well as the personnel operating them. Variations between operators and differences between successive readings taken at the same point on the wire are often large enough to nullify an entire set of data. The solution to this measurement problem has been to disregard all individual readings and to base the results on statistical samples. A standard wire has been selected and its diameter ascertained by averaging a series of measurements made on an electron microscope. Fig. 2 illustrates the method of measuring a new specimen. The readings are taken on a 625-power monocular microscope equipped with a Filar eyepiece. A series of measurements is made on the standard wire and another series on the specimen. The difference between process averages, expressed as "Filar divisions," is converted into microinches and is added to or subtracted from the accepted diameter of the standard to give the diameter of the specimen. Acceptance or rejection of wire is on the basis of meeting limits of 0.000290 inch ±0.000010 inch for the diameter.

Electron Tube Performance

A steady flow of normally distributed tube parts on the manufacturing line has permitted production of tubes that satisfy the design objectives. Since the characteristic of particular importance to the transmission system is transconductance, this parameter is required to meet strict limits in regard to the process average. The limits which are being met are given in Table III.

The results achieved in control of transconductance are illustrated in Fig. 3. This is an actual production control chart for the *W.E. 435A* tube.

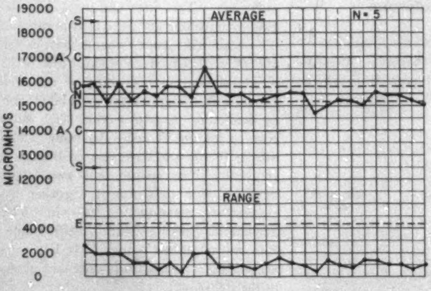


Fig. 3. Control chart for transconductance of W.E. 435A electron tube

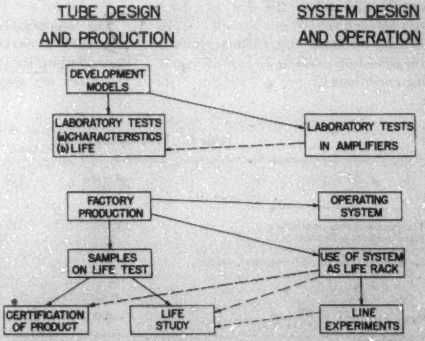

TUBE DESIGN AND PRODUCTION
SYSTEM DESIGN AND OPERATION
DEVELOPMENT MODELS
LABORATORY TESTS (a) CHARACTERISTICS (b) LIFE
LABORATORY TESTS IN AMPLIFIERS
FACTORY PRODUCTION
OPERATING SYSTEM
SAMPLES ON LIFE TEST
USE OF SYSTEM AS LIFE RACK
CERTIFICATION OF PRODUCT
LIFE STUDY
LINE EXPERIMENTS

Fig. 4 (right). Flow diagram for study of tube life

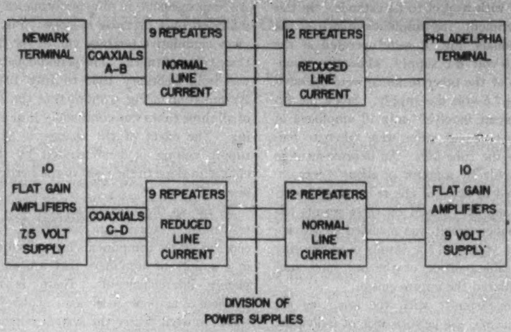


Fig. 5. Line experiments on the Newark-Philadelphia L3 system

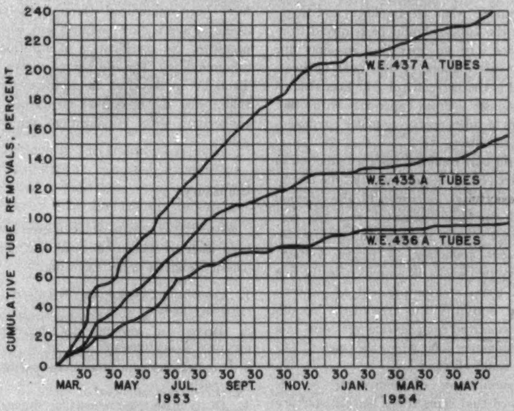


Fig. 6. Cumulative tube removals from the Newark-Philadelphia L3 system

Several sets of limits are indicated on the chart where N is the nominal (desired objective) value and A is half of the permissible range for individuals:

1. The test specification limits applicable to individual tubes: S limits $=N\pm A$.
2. The C limits for controlling the sample average X': C limits $=N\pm 0.5\ A$.
3. The D limits for controlling the process average X'': D limits $=N\pm 0.1\ A$.
4. The E limit or control limit for range R: E limit $=1.48\ A$.

To satisfy the system requirements, random shipment of tubes is permitted only if the following eligibility requirements are met:

1. X' falls within the C limits.
2. X' falls outside the C limits but the six preceding consecutive averages fall within the C limits.
3. X' and the six preceding consecutive averages do not fall outside the same D limit. (This in effect controls X''.)
4. The range R falls within the E limit, or falls outside the E limit but the six preceding consecutive ranges fall within the E limit.

Loss of eligibility requires shipment of product in small groups to a forced symmetrical distribution. The effectiveness of the control measures is demonstrated by the fact that, for all but a short period of time, eligibility for random shipment of tubes has been maintained.

When it is desirable to control more than one characteristic of the product, the more important one is subject to control with restrictions on shipment. The other characteristics are monitored by a records method where plotting control charts is mandatory, but shipment of the product is not contingent on the evidence of the charts. However, the producer is responsible for bringing the process under control. A detailed discussion of these quality control requirements has been published previously.[1]

Table IV. Reliability of Tubes in Entire System

Characteristic	Objective	Experience
Sudden failure rate	<1 per 1,000 sockets per month	approximately 2 per 1,000
Tube life, hours	15,000 average	
W.E. 435 A		18,000
W.E. 436 A		>30,000
W.E. 437 A		10,000

Tube and System Performance

The life of electron tubes is, figuratively speaking, influenced by both heredity and environment. The heredity of the electron tubes depends on the processes at the factory; the environment is furnished by the field operating conditions. The tube life problem cannot be solved at the factory alone but requires co-operation with the system designer and extensive analysis of performance data in the field. Furthermore, co-operation should be continuous and should start in the design stage.

The life and reliability problems of the *L-3* system are co-ordinated by a committee comprising representatives of tube design, tube manufacture, circuit design, circuit manufacture, and system operation (the customer). The function of the committee is to act as a planning board and to review and disseminate information. The detailed work is done in smaller subcommittees.

A flow diagram showing the approach to the life problems is given in Fig. 4. The left half of the chart shows the functions of the tube designers and manufacturers while the right half lists the system design and operation effort. Solid arrows indicate the flow of product, and dotted arrows the flow of information. It will be noted that, while tubes were in the development stage, models were submitted for laboratory amplifier testing. Later, when full production had been established, the system was used as a life rack and correlation studies were made with data obtained on life samples at the tube factory.

Tube identification has been maintained by designating weekly production lots and by numbering individual tubes. Weekly lots are kept as homogeneous as possible in regard to materials and processes. The culmination of this segregation program is the certification of product by these weekly production lots. To assure a temporary field life objective of at least 8,000 hours, tubes are accepted for system use if life samples successfully pass a 2,000-hour test. Where the sample fails to meet the 2,000-hour criterion, the lot is withheld from system use. The factory life sample (five tubes from lots averaging about 125 tubes) is augmented by data from the field correlation studies mentioned in the foregoings. In other words, a portion of the system is used as a life rack to provide data for certification of the tubes intended for general system use.

The initial installation of equipment was in a field trial between Newark and Philadelphia. The experience of this trial in which the entire circuit was used as a life rack (handling about 500 tubes) showed that the life picture was very uneven and was related to tube lots. This furnished some valuable clues as to where to look for sources of trouble and emphasized the benefits of the committee operation.

During this phase of operation two designed experiments were initiated to ascertain the effects on tube life of different operating conditions. The experiments were run simultaneously and are shown schematically in Fig. 5. The system comprises two sets of coaxial cables, designated *A-B* and *C-D* in Fig. 5, and is divided so that approximately equal sections are powered from each of the two terminals (Newark and Philadelphia). In two sections the line current was reduced to 95 per cent of the normal operating value and was left at normal in the other two sections. The effect was operation of half of the tubes at reduced heater power and the remainder at normal heater power. The experiment was balanced to permit statistical treatment of the data from coaxials and power sections. While about 15-per-cent improvement in average tube life was noted at the reduced power, this was not considered sufficient to warrant a change and the system continues to operate at the original power level.

The second of the designed experiments evaluated tube life in relation to d-c feedback. In the amplifier the control grid of each tube was returned to a +9-volt potential rather than to ground, and about 10 volts were developed across each cathode resistor. As a result of this combination bias, the net bias of each control grid is in the order of −1 volt with respect to its cathode. In this experiment, the amplifiers at one of the terminals, Fig. 5, were operated at the +9.0-volt d-c supply, while the amplifiers at the other terminal were operated at +7.5-volt d-c supply. Since the experiment involved only 10 amplifiers at each terminal, steps were taken to balance the tube lots. An improvement in tube life by a factor of about 3 was observed in favor of the reduced grid voltage supply. Based on these results the amplifier circuit was modified to provide +7.5 volts and the results of monitoring a sizable portion of the line have substantiated the improvement.

Concurrently with the foregoing experiments, the performance of individual tubes was analyzed. This dealt with the measurements being used in the field to determine when an amplifier, and therefore the tubes, should be removed from service. In an operating amplifier the only tube parameter which can be measured is the net control grid bias defined previously. From experinece with other types of tubes operating under combined-bias conditions, it is known that the change in grid bias resulting from a reduction in heater voltage correlates with tube thermionic activity. Accordingly, the criterion used for tube removal was the percentage change in bias caused by a 10-per-cent reduction of heater voltage. Every two weeks all amplifiers in the line were tested and the per-cent ΔE_c (thermionic activity) was recorded. Amplifiers removed from service were sent to a maintenance point where the per-cent ΔE_c was again checked.

Examination of individual records showed that some tubes had satisfactory activity readings in the field but failed to meet maintenance limits. On-the-spot investigation and correlation studies disclosed that this discrepancy was only apparent and was caused by differences in measuring equipment and supply voltages. A graded pattern of limits has been established to compensate for these differences.

System Reliability

A simple and effective way of showing the field life performance of tubes is a plot of cumulative removals against calender time. In a stable system, i.e,. when the replacement tubes come from essentially the same universe as the tubes already in service and the environment is constant, the plot will be a straight line. An improvement in life performance is indicated by a decrease in slope. Fig. 6 gives cumulative removal curves for the three types of tubes. The period covered was from February 1953 to July 1954. Examination of Fig. 6 shows that the life of all three tubes was continually improving. The effect of the change in grid supply voltage is demonstrated by the change in slope to the right of the vertical line.

As far as the broader aspects of the transmission system are concerned, the controls on the components have resulted in a supply of amplifiers that are completely interchangeable. There is no problem of amplifier selection. Table IV shows to what degree the system reliability objectives have been met in regard to tube performance. The sudden failures have been minimized through factory assembly and testing measures and, where necessary, by specification of additional requirements. The life figures shown in Table IV are based on end-of-life criteria that insure the objectives mentioned in connection with equalization requirements.

Summary

A variety of statistical techniques have been used to achieve the results to date. These are given in the following:

1. Probability theory for derivation of reliability limits.
2. Analysis of variance for measuring accuracy of instruments.
3. Control charts for production control of tube parts.
4. Distribution requirements for shipment of product.
5. Correlation studies for field and maintenance test limits.
6. Designed experiments for field life evaluation.

While there is more to be done to improve some of the reliability factors, it is felt that the results achieved could not have been accomplished without the use of these statistical tools. In addition, the effective operation of the co-ordinating committee is credited with a share in this achievement.

Reference

1. THE L3 COAXIAL SYSTEM QUALITY CONTROL REQUIREMENTS, H. F. Dodge, B. J. Kinsburg, M. K. Kruger. *Bell System Technical Journal*, New York, N. Y., vol. 32, Jul. 1953, pp. 943–67.

Delay Distortion Correction

W. D. CANNON
FELLOW AIEE

IN FACSIMILE and data transmission systems the signals are largely transient in nature and are usually in the form of a modulated wave for transmission purposes. The modulating signals may be in the form of isolated pulses, periodic waves, or various other wave shapes depending upon the nature of the communication device or system. Modulated pulse envelopes are thus present in the transmission medium in various combinations of amplitude, duration, and relative positions. It is therefore necessary to consider the transient response characteristics of transmission systems and associated networks as well as their steady-state properties.

A fundamental problem in the transmission of modulated pulses is the resulting distortion arising from imperfections in the transmission medium and circuitry in the form of amplitude and delay variations with respect to frequency and bandwidth limitations. In transmission the pulse envelopes may be broadened, displaced from their true positions, or destroyed beyond recognition. Amplitude distortion usually can be corrected by simple means. Delay distortion correction, on the other hand, involves relatively more complicated measuring techniques and compensating networks.

It is usual to provide correction for envelope delay distortion and ignore other forms of delay distortion. An instrument has been developed for the measurement of envelope delay characteristics as well as amplitude characteristics.[1] By the use of this instrument these characteristics may be displayed continuously by a cathode-ray oscilloscope. Thus each alteration in the characteristic by the insertion of corrective measures is immediately displayed so that the process of corrective insertion and observation is continued until the characteristic is within a prescribed tolerance. The usual passive recurrent-structure corrective network is impractical for the rapid insertion of new delay values in small steps for observation. This paper includes consideration of distortion properties and the development of a more practical means of correction, particularly for use in conjunction with the envelope delay measuring instrument.

Pulse Transmission Characteristics

Since facsimile and data signals contain wave shapes in the form of pulses and depend upon the relative phases of the component frequencies as well as upon relative amplitudes for preservation of wave shape, it is necessary to consider both amplitude and phase responses. It is usual to predict the transient response of a circuit or network from the steady-state amplitude and phase responses. These responses may be expressed in closed form by relations such as a Fourier integral equation:

$$A(t)=\frac{1}{2}S_0+\frac{1}{\pi}\int_0^\infty \frac{S(\omega)}{\omega}\sin\,[\omega t+\varphi(\omega)]\,d\omega \tag{1}$$

where $A(t)$ is response due to unit step function, S_0 is response after the transient has disappeared, $S(\omega)$ is amplitude response as a function of ω, and $\varphi(\omega)$ is phase response as a function of ω.

The phase response $\varphi(\omega)$ can in general be expressed as the sum of three components. The first is the minimum phase component which has a definite relationship to the amplitude response and is therefore of particular interest with respect to limited bandwidths. In minimum phase systems equation 1 can be expressed in terms of either the amplitude response $S(\omega)$ or the phase response $\varphi(\omega)$. The second is a linear component which represents a constant transmission time for all frequencies. The third component is independent of the amplitude response and hence may be used for application in networks for correcting the phase response to be a linear characteristic with respect to frequency. In this case $S(\omega)$ is a constant and transient response $A(t)$ is dependent upon the phase response $\varphi(\omega)$. These components give rise to different forms of phase distortion. The properties of some of these forms will be considered briefly.

The amplitude response $S(\omega)$ and the phase response $\varphi(\omega)$ are usually considered in terms of the steady-state parameters α, attenuation constant; and β, the phase shift constant so that the propagation constant is $\alpha+j\beta$. The phase-related parameters that are associated with facsimile signal distortion may be expressed in terms β/ω which is the phase delay and $d\beta/d\omega$ which is the envelope or group delay. In addition, higher order derivatives are sometimes of importance particularly when aperiodic waves are involved.[2, 3]

In the transmission of periodic waves or pulse trains such as are generated when a number of equally spaced lines are scanned, it is convenient to express such waves in the form of a Fourier series

$$\begin{aligned}E(t)=\frac{a_0}{2}+a_1\sin pt+a_2\sin 2pt+\\ \ldots a_n\sin npt+b_1\cos pt+\\ b_2\cos 2pt+\ldots b_n\cos npt\end{aligned} \tag{2}$$

where the parameters may be determined by the usual Fourier methods. If used to modulate a carrier, double side band, the resulting transmission is

$$V_1=A_1\cos\omega_c t[1+ME(t)] \tag{3}$$

where $\omega_c/2\pi$ is carrier frequency, A_1 is carrier amplitude, and M is modulation factor. If this wave V_1 is transmitted through a medium having a linear phase shift as shown in Fig. 1, curve 1, each sinusoidal component of $E(t)$ will be received in the form

$$\begin{aligned}V_2=A\cos\omega_c(t+k)+\frac{Ab_nM}{2}\cos(\omega_c+\\ np)(t+k)+\frac{Ab_nM}{2}\cos(\omega_c-np)(t+k)\end{aligned} \tag{4}$$

Hence, the carrier and two side-band frequencies will be delayed in transmission by the factor $\beta/\omega=d\beta/d\omega=k$ so that all components of $E(t)$ will be delayed equally and the wave shape will be preserved in transmission. If the medium has a linear phase shift which does not intercept the β-axis at 0 or $m\pi$, where m is an integer, as shown by curve 2, Fig. 1, a form of distortion occurs known as linear phase intercept distortion. Here the phase delay $\beta/\omega=k+\beta_1/\omega$ and the envelope delay $d\beta/d\omega=k$ are not equal. In this case each sinusoidal component of $E(t)$ will be received in the form

---

Paper 56-107, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing November 29, 1955.

W. D. CANNON is with the Western Union Telegraph Company, New York, N. Y.

The author acknowledges the guidance and assistance of many people concerned with the measurement and correction of delay distortion, especially Henry F. Burkhard and Chester S. Friedman of the U. S. Army Signal Corps, and F. B. Bramhall, Albert Boggs, and D. J. Bertuccio of the Western Union Telegraph Company.

The development of the instrument described in this paper was one phase of a comprehensive investigation of delay distortion and correction conducted by the Western Union development and research organization under U. S. Army Signal Corps sponsorship.

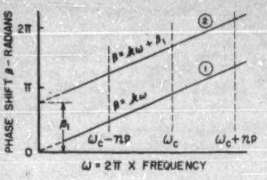


Fig. 1. Linear phase shift characteristics

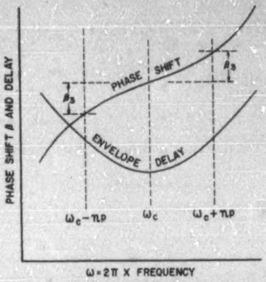


Fig. 3. Symmetrical phase and envelope delay characteristics

Table I. Active Networks

Group	Source Impedance	Source Voltage	Load Impedance	Load Voltage
A	Low	Transverse	High	Transverse
B	High	Transverse	Low	Transverse
C	Low	Transverse	High	Longitudinal
D	High	Transverse	Low	Longitudinal
E	Low	Longitudinal	High	Longitudinal
F	High	Longitudinal	Low	Longitudinal
G	Low	Longitudinal	High	Transverse
H	High	Longitudinal	Low	Transverse

$$V_2 = A\cos[\omega_c(t+k)+\beta_1] + \frac{Ab_nM}{2}\times \cos[(\omega_c+np)(t+k)+\beta_1] + \frac{Ab_nM}{2}\cos[(\omega_c-np)(t+k)+\beta_1] \quad (5)$$

which in envelope form is

$$V_2 = A\cos[\omega_c(t+k)+\beta_1][1+Mb_n\times \cos np(t+k)] \quad (6)$$

Hence the carrier is delayed by the phase delay and the envelope is delayed by the envelope delay of the medium.

The question often arises as to whether linear phase intercept distortion is of importance in facsimile and data transmission systems where information may be present in the form of modulated pulses, whereas unmodulated signals such as the $E(t)$ waves of equation 1 when transmitted directly over a circuit could be distorted beyond recognition unless the intercept is 0 or $m\pi$. From the foregoing it can be seen that in the case of modulated signals the envelope is not distorted in transmission but that the carrier is shifted in phase under the envelope. Hence this form of distortion may be neglected, unless some use is to be made of the individual cycles of the carrier, in which case correction need be made only for envelope delay. However, the peaks of the carrier cycles are often used for marking or sampling rather than the recovered envelope. In this case the necessity for intercept correction is minimized if the carrier frequency is high relative to the modulation frequency since the constant carrier phase shift angle β_1 will then represent a small time displacement relative to the signal interval. This relationship may be effected, if it should not naturally exist in the system, by frequency multiplication at the receiver which will multiply the carrier frequency ω_c and leave the modulation frequency unchanged.

For a medium in which the phase shift is not linear, a number of different types of distortion arise depending upon the phase characteristic. A few types of phase characteristics are discussed in the following. Since a delay common to all signal frequencies produces no wave-form distortion, that part of the phase shift which is linear with frequency may be ignored for present purposes. The results of nonlinear distortion may for convenience be expressed with respect to the average value of the phase shifts β_4 and β_5 of the side frequencies of a component of $E(t)$, Fig 2. The received modulated component then will be

$$V_3 = A\cos(\omega_c t+\beta_2) + \frac{Ab_nM}{2}\times \cos[(\omega_c+np)t+\beta_3] + \frac{Ab_nM}{2}\cos[(\omega_c-np)t-\beta_3] \quad (7)$$

$$= A\cos\omega_c t[\cos\beta_2 + b_nM\cos(npt+\beta_3)] - A\sin\omega_c t\sin\beta_2 \quad (8)$$

where

$$\beta_2 = \frac{\beta_5-\beta_4}{2} \text{ and } \beta_3 = \frac{\beta_4+\beta_5}{2} \quad (9)$$

which in envelope form becomes

$$V_3 = A\sqrt{1+2b_nM\cos\beta_2\cos(npt+\beta_3)+b_n^2M^2\cos^2(npt+\beta_3)}\times \cos(\omega_c t+\theta) \quad (10)$$

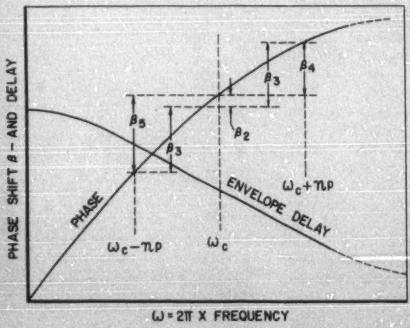


Fig. 2 (left). Nonlinear phase shift characteristics

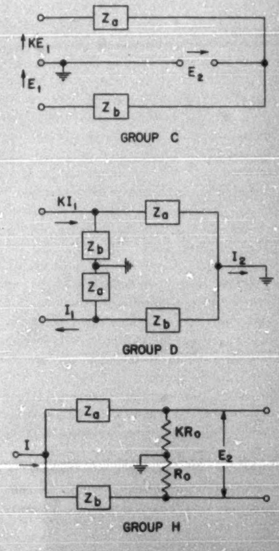


Fig. 4 (right). Typical network configurations for groups C, D, and H

Vol. 7
Pt. 1
1956

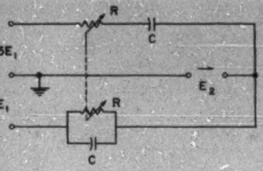


Fig. 5 (left). Phase shifting K=5 network applied to group C

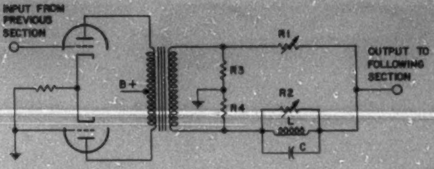


Fig. 6 (right). Active corrective network section

where

$$\theta=\tan^{-1}\frac{-\sin\beta_2}{\cos\beta_2+b_nM\cos(npt+\beta_2)} \qquad (11)$$

It will be observed that there is envelope distortion due to the presence of the cos β_2 term and some degree of phase modulation due to the variation of θ with npt. If the phase shift $\beta_2=\pi/2$ the envelope of equation 10 becomes

$$A\sqrt{1+b_n^2M^2\cos^2(npt+\beta_2)} \qquad (12)$$

$$=A\left[1+\frac{b_n^2M^2}{2}\cos^2(npt+\beta_2)-\frac{b_n^4M^4}{8}\cos^4(npt+\beta_2)\right]\ldots$$

The first term of this series becomes

$$A\left[1+\frac{b_n^2M^2}{4}+\frac{b_n^2M^2}{4}\cos 2(npt+\beta_2)\right] \qquad (13)$$

Hence the original modulation is destroyed and a double-frequency term of the modulation as well as decreasing amounts of higher even order terms arise. If phase shift β_2 is 0 or $m\pi$ the envelope is

$$A\sqrt{1\pm 2b_nM\cos(npt+\beta_2)+b_n^2M^2\cos^2(npt+\beta_2)} \qquad (14)$$

$$=A[1\pm b_nM\cos(npt+\beta_2)]$$

whence there is no envelope distortion but merely a reversal of sign for even values of m which is of no consequence. The result for double side-band transmission is thus analogous to intercept distortion with respect to the intercept points on the vertical axis which has been modulated up to carrier position and has the same relation in phase shift of $\beta_2=0$ or $m\pi$.

If the carrier is located at a point of odd symmetry on the phase characteristic as shown in Fig. 3, β_2 will be zero and from equation 8 the received component will be

$$V_2=A\cos\omega_c t[1+b_nM\cos(npt+\beta_2)] \qquad (15)$$

and no intercept type of distortion occurs. However, the different components of $E(t)$ will not be delayed equally and some envelope distortion will remain. It is thus desirable in facsimile double sideband transmission practice to locate the carrier at a point of even symmetry on the delay characteristic in order to minimize signal distortion.

In the foregoing it has been assumed that the amplitude characteristic is constant throughout the useful frequency range. If both amplitude and delay characteristics are not constant with frequency other forms of distortion take place. One particular type takes place in the form of echoes which in recorded form results in ghosts. An echo results if the amplitude has a cosine variation and the phase characteristic has a corresponding sine variation, or conversely, with respect to frequency.[4,5] In this case the spacing of the echo from the original pulse depends upon the spacing of these sinusoidal variations with respect to frequency.

However, it is usually sufficient to make correction only for envelope delay distortion in addition to the usual amplitude correction. It is not necessary to consider phase shift directly as such. When satisfactory correction has been made for envelope delay distortion, it usually will be found that phase delay distortion is within tolerable limits particularly if a sufficient number of carrier cycles exist under the modulation envelope. Accordingly, networks have been considered here mainly for the correction of envelope delay distortion.

Corrective Network Characteristics

For the compensation of delay distortion in carrier "vehicle" bands and associated equipment, types of corrective networks which may be readily designed and individually adjusted to meet the conditions of a particular application are required. For use with the delay measuring instrument, they must be readily adjustable either continuously or in small steps to compensate for a deficiency at any point in the useful frequency range. Such networks, generally designated as all-pass are usually designed to have an amplitude characteristic which is essentially independent of frequency throughout the useful frequency range These networks may be passive or active. Active networks have a property not possessed by passive networks that the delay characteristic can be readily varied over a wide range, whereas passive networks that permit the delay characteristic to be varied over the necessary range become impracticably complicated. Also, it is the property of some types of active networks that in addition to the adjustable delay feature, the amplitude characteristics may be varied over a limited range to compensate circuit amplitude distortion or amplitude distortion introduced by other networks.

The passive networks usually considered for distortion correction are the constant-resistance recurrent structures designed to match the line impedance.[6] Active networks may be designed for various conditions of input and output terminating impedances and methods of injecting the input and extracting the output. A number of these are given in Table I.

It can be readily shown that the transfer function of a delay network having an attenuation or gain which is constant with frequency can be expressed in the form

$$N\frac{A\pm jB}{A\mp jB}=N\underline{/\beta} \qquad (16)$$

or other equivalent expressions, where B is $2\tan^{-1}\pm B/A$ and N is a constant.

A and B are functions of ω and the network parameters. For the present purposes the constant N may be neglected.

In general the groups listed in the table can be realized by a number of network configurations. A typical network configuration for groups C, D, and H is shown in Fig. 4 for which the output to input transfer ratio is of the form

$$\frac{E_2}{E_1}=\frac{KZ_b-Z_a}{Z_a+Z_b} \qquad (17)$$

where Z_a and Z_b are functions of ω and the network parameters, and K is a proportionality factor. The proportionality factor K may be real or complex. However, only real numbers will be considered here.

Under certain prescribed relations among the parameters, the factor K may take on integer values. If $K=1$ equation 17 is directly in the form of equation 16. The result is the well-known delay network in which the conditions are satisfied by a resistance for Z_a and a reactance for Z_b so that

Fig. 7. Passive corrective network

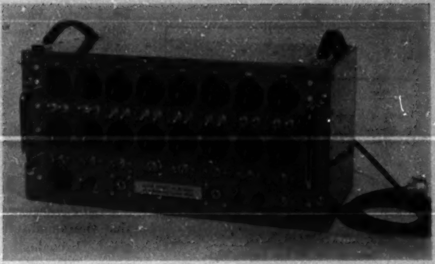


Fig. 8. Active corrective network

$$\frac{E_2}{E_1}=\frac{jX-R}{jX+R} \tag{18}$$

and

$$\beta=2\tan^{-1}\frac{X}{R} \tag{19}$$

If $K=3$ the conditions can be satisfied by the use of a resistance for Z_a and a resistance and reactance in parallel for Z_b so that

$$\frac{E_2}{E_1}=\frac{2jX-R}{2jX+R} \tag{20}$$

and

$$\beta=2\tan^{-1}\frac{2X}{R} \tag{21}$$

For $K=5$, the conditions can be satisfied by impedance functions for Z_a and Z_b such as

$$Z_a=R_1+jX \text{ and } Z_b=\frac{jMR_2X}{R_2+jMX} \tag{22}$$

where M is a constant.

However, a very useful form occurs when $M=1$ and $R_1=R_2=R$ in which the output-input transfer ratio is

$$\frac{E_2}{E_1}=\frac{3jRX-(R^2-X^2)}{3jRX+(R^2-X^2)} \tag{23}$$

and

$$\beta=2\tan^{-1}\frac{3RX}{R^2-X^2} \tag{24}$$

The $K=3$ and $K=5$ series of networks are particularly useful in phase shifting and delay networks. A very simple phase shifting network applied to group C is shown in Fig. 5 in which, from equation 24,

$$\beta=2\tan^{-1}\frac{3\omega RC}{\omega^2R^2C^2-1} \tag{25}$$

A range for fixed values of ω and C from approximately 0 to 360 degrees is obtained with constant amplitude throughout this range merely by simultaneously varying the two equal resistors R.

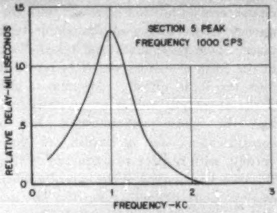


Fig. 9. Envelope delay characteristic of a passive network section

The $K=3$ series has been selected for delay distortion correction where variable delay properties are essential. The schematic arrangement of a section is shown in Fig. 6 where the resistor R_1 is used for Z_a and the parallel arrangement of resistor R_2, inductor L and capacitor C is used for Z_b. The $K=3$ factor is obtained by the divider R_3 and R_4 such that $R_3/R_4=3$. The delay characteristic is varied by R_1 and R_2. If these resistors are maintained equal, the delay characteristic will be dependent upon their common value while the amplitude characteristic will remain constant for all settings and from equation 21 β will be

$$\beta=2\tan^{-1}\frac{2\frac{\omega}{\omega_0}\sqrt{L/C}}{\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]R} \tag{26}$$

If, in addition, it is desirable to produce a variable amplitude characteristic, this can be accomplished by offsetting resistor R_2 with respect to resistor R_1. In this case the output-input transfer ratio and delay can readily be shown to be

$$\frac{E_2}{E_1}=\frac{jX(3R_2-R_1)-R_1R_2}{jX(R_2+R_1)+R_1R_2} \tag{27}$$

$$\beta=\tan^{-1}\frac{X}{\frac{R_1R_2}{3R_2-R_1}}+\tan^{-1}\frac{X}{\frac{R_1R_2}{R_1+R_2}} \tag{28}$$

where

$$X=\frac{\frac{\omega}{\omega_0}\sqrt{L/C}}{1-\left(\frac{\omega}{\omega_0}\right)^2} \tag{29}$$

If R_2 is larger than R_1 the amplitude characteristic, from equation 27, will be peaked upward in the vicinity of the resonant frequency ω_0. If R_2 is smaller than R_1 it will be peaked downward. Thus with these networks it is possible to adjust the delay characteristic over a wide range and the amplitude characteristic over a somewhat restricted range essentially independent of each other merely by adjusting two resistors and maintaining a proper relationship between them in accordance with equations 27 and 28. Practically, it is usually necessary to change only R_2 from equality with R_1 to change from a flat amplitude characteristic to a peaked characteristic in the vicinity of ω_0. This is not possible with passive networks which usually require separate networks for amplitude and delay distortion correction since each alteration of the amplitude characteristic of a purely amplitude corrective network introduces some delay distortion which in turn must be offset in the delay corrective network. For this reason, amplitude distortion correction is usually made first. With the $K=3$ active network, this process is inherently accomplished.

Application of Corrective Networks

A great variety of delay curve shapes may be encountered among the transmission circuits to be equalized with corrective networks. Not only are there large variations among the different types of circuits, but even the same type requires different amounts of correction throughout the essential frequency range. For any circuit to be corrected, the desired result to be achieved is a uniform amplitude and delay response flat within

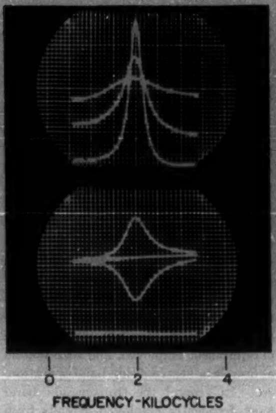


Fig. 10. Typical envelope delay and amplitude characteristics of an active network section

Top—Relative delay, 100 microseconds, per small divisions
Bottom—Amplitude variation
Curve data apply also to Figs. 11 and 12 and Figs. 14–16

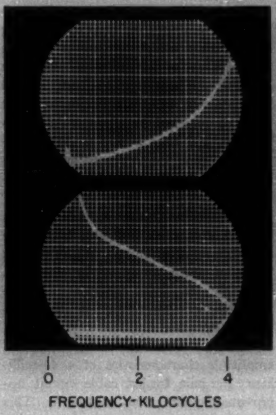


Fig. 11. Delay and amplitude characteristics of a 44-millihenry-per-mile loaded cable—uncorrected

prescribed limits over the essential band width. Since delay cannot be subtracted from the circuit, the problem resolves itself into one of raising the delay of all points to the vicinity of the highest delay value on the characteristic within the essential band width.

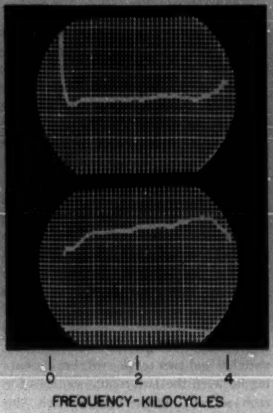


Fig. 12. Delay and amplitude characteristics of a 44-millihenry-per-mile loaded cable—corrected

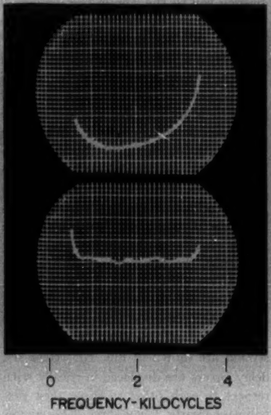


Fig. 13. Delay characteristics of a radio beam circuit—uncorrected and corrected

Relative delay, 100 microseconds, per small divisions

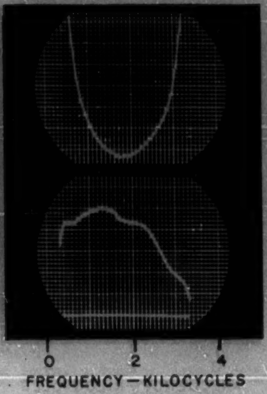


Fig. 14. Delay and amplitude characteristics of a type-K carrier circuit—uncorrected

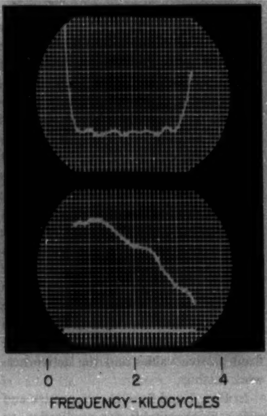


Fig. 15. Delay and amplitude characteristics of a type-K carrier circuit—delay corrected, amplitude uncorrected

The application of both passive and active networks is particularly advantageous in conjunction with the delay measuring instrument by which the delay characteristic is continuously displayed. The passive sections possessing relatively large lumped delay values, preferably are first inserted for a rough correction. After this has been done, the remaining correction may be accomplished with active network sections, and those areas filled in where delay deficiencies exist. The ability to vary the delay of each individual section allows much finer adjustments to be made than is practical with passive sections. If, after obtaining a refined delay correction, amplitude distortion exists as a result of insufficient amplitude correction, or amplitude distortion inserted by the passive sections, the amplitude characteristic may be corrected with some limitations by offsetting the adjustable resistors of the active network sections as described in the foregoing.

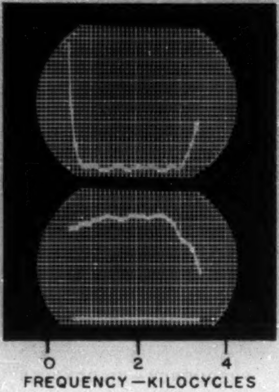


Fig. 16. Delay and amplitude characteristics of a type-K carrier circuit—corrected

In some applications such as in military circuits, it might be desirable to provide packaged units containing a number of passive sections and other units containing a number of active sections. In this case the units would be left in circuit as adjusted and would be available for making a quick change in case the circuit makeup is altered. In other cases this procedure would not be economical. The preferred method then would be to insert other passive and active sections in the circuit corresponding to the sections in use in the packaged units, or specially select or design passive sections which simulate the combined delay of the passive and active sections. The design parameters can be obtained from the fixed passive values and the active network settings.

In Figs. 7 and 8 are shown passive and active units suitable for application to facsimile circuits as discussed in the foregoing. The passive unit consists of 25 bridged-T sections and an input and output transformer. Each section has a fixed delay distribution and an attenuation at the frequency of peak delay ranging from about 0.8 decibel in the low-frequency sections to about 1 decibel in the upper frequency sections. The peak delay occurs at selected frequencies ranging from 600 cycles for section 1 to 3,000 cycles for section 25. A representative curve of the delay distribution of one section is shown in Fig. 9. The sections are arranged in four groups to simulate assigned delay distributions. Switches are provided to permit any group or any combination of sections to be inserted.

The active network unit consists of eight type $K=3$ sections as shown in Fig. 8 with electron tube coupling between sections, an input gain control, and a twin-triode output amplifier stage. The peak delay values of the sections occur at approximately 375 cycles per second intervals from approximately 600 cycles to 3,200 cycles, with slightly closer spacing around the center of the band. As with the passive network, each section may be individually inserted or omitted for any circuit correction. Toggle switches are provided on the panel of the unit for this purpose. In addition to providing individual section selection, this unit has the additional advantage of individually variable delay for each section. The delay is varied with two rotary switches and two toggle switches appearing between the two rotary switches for each section. With these controls the peak delay of any section may be varied in 22 steps from a minimum of about 100 microseconds to maximum of about 1,500 microseconds. When the sections are inserted in tandem, the delay at any frequency will be the sum of the individual section delays at that frequency. The amount of correction inserted may range from about 100 microseconds when only a few sections are in circuit to approximately 3,000 microseconds when all sections are included and set at their maximum peak delay values. When it is desirable to change only the delay distribution, the switches are set in pairs with the switches of each section at the same switch position. If in addition it is desirable to make an amplitude correction in the vicinity of the section peak delay frequency, the lower rotary switch of a pair is set forward for an increase in amplitude and backward for a decrease in amplitude. Representative curves of the delay and amplitude distributions of one section are shown in Fig. 10. Here the delay curves are displaced vertically with respect to each other to show their distribution clearly. The three delay curves depict settings as provided in the active unit near the minimum, intermediate, and maximum peak delay values. The amplitude curves correspond to the intermediate delay distribution for a flat amplitude characteristic, and for intermediate amplitudes peaked upward and downward.

In addition to their use in effecting a particular delay distribution at constant amplitude, these units also have the property that various amplitude distributions can be simulated at an essentially constant delay. For example, by proper switch settings a vestigial filter amplitude characteristic can be simulated at constant delay throughout the effective band width.

Results of Distortion Correction

Facsimile and data processing equipment are required to operate satisfactorily over various types of commercial and military circuits. and various combinations of these circuits. These transmission facilities consist of various physical, carrier, and radio circuits having a wide range of envelope delay distortion and often an appreciable amount of amplitude distortion. The over-all frequency range necessary to provide satisfactory operation depends upon the type of equipment employed, and such factors as double or vestigial side-band operation, etc. In general the transmission band most frequently employed extends from about 600 or 800 cycles to about 3,000 cycles for the most important frequencies contained in the signals. Frequencies involved in transmission near the carrier frequency are usually of greater importance than frequencies remote from the carrier frequency.

For application to facsimile circuits, a number of corrective sections are connected in tandem, usually at the receiving end of the circuit; the number depends upon the type of delay characteristic to be corrected. These sections are spaced at frequency intervals throughout the band as required to fill in a delay deficiency, and each active section is adjusted so that the combined delay of all sections is the desired delay compensating characteristic. In general, relatively broad individual curves spaced far apart are required when the delay to be corrected is small in value. Where a large value of delay needs to be corrected, relatively steep delay curves can be used with the sections spaced at small requency intervals

In Fig. 11 are shown the delay and amplitude characteristics of a 44-millihenry-per-mile loaded cable circuit approximately 125 miles in length. This circuit was corrected by use of the active network alone with the result as shown in Fig. 12 in which delay characteristics is flat within the usual allowable tolerance of ±250 microseconds from approximately 500 to 4,000 cycles. A rough amplitude correction was obtained by use of the active network as shown by the lower traces of these oscillograms. It should be noticed that since amplitude can be either added or subtracted, in the corrective process the amplitude was decreased for the lower frequencies and increased for the upper frequencies by the active networks. Likewise, in Fig. 13 are shown results as obtained on a radio beam circuit approximately 200 miles in length. The upper curve of this oscillogram is the delay characteristic without correction and the lower curve is the delay characteristic after correction has been applied. The corrective process is well illustrated by the oscillograms of Figs. 14–16, which utilize a long type-K carrier circuit. Fig. 14 shows the uncorrected delay and amplitude characteristics; in Fig. 15, the delay characteristic has been corrected; and Fig. 16, a partial amplitude correction has been obtained. It should be pointed out that in making the amplitude correction, the delay characteristic was unaffected and, for all practical purposes, the delay and amplitude corrections can be made independently of each other.

These networks provide a readily adjustable means for obtaining various delay characteristic shapes in addition to those normally present in facsimile circuits for the appraisal of facsimile recorded copy with respect to delay distortion. A large number of such test conditions have been investigated which comprise various tolerances from a mean value. For the loaded physical circuit, Figs. 11 and 12, facsimile operation is considerably improved by distortion correction. For type-K circuits distortion correction may give sufficient improvement to permit operation at double speed with vestigial operation as compared to double side-band operation with the carrier near a point of even symmetry on the delay characteristic. Also, delay distortion correction may permit splitting the normal voice band circuit into two subbands so that normal operating speed can be obtained in each of these bands by use of vestigial operation. In general, the appraisal of the various factors involved indicate that the tolerance limit for satisfactory facsimile operation is of the order of ±400 microseconds from the mean value per 1,000 cycles of modulation (inverse relationship), whereas the usual quoted value is generally ±250 microseconds.

References

1. AN ENVELOPE DELAY MEASURING INSTRUMENT IN THE AUDIO FREQUENCY RANGE, W. D. Cannon. *AIEE Transactions*, vol. 74, pt. I, Jan. 1956, pp. 710–17.

2. THE BUILDING UP OF SINUSOIDAL CURRENTS IN LONG PERIODICALLY LOADED LINES, John R. Carson. *Bell System Technical Journal*, New York, N. Y., Oct. 1924, pp. 558–66.

3. PHASE DISTORTION AND PHASE DISTORTION CORRECTION, Sallie P. Meade. *Ibid.*, May 1928, pp. 195–224.

4. THE INTERPRETATION OF AMPLITUDE AND PHASE DISTORTION IN TERMS OF PAIRED ECHOES, H. A. Wheeler. *Proceedings*, Institute of Radio Engineers, New York, N. Y., June 1939, pp. 359–85.

5. INFLUENCE OF ECHOES ON TELEVISION TRANSMISSION, P. Mertz. *Journal*, Society of Motion Picture and Television Engineers, Easton, Pa., May 1953, pp. 572–96.

6. DISTORTION CORRECTION IN ELECTRICAL CIRCUITS WITH CONSTANT-RESISTANCE RECURRENT NETWORKS, Otto J. Zobel. *Bell System Technical Journal*, New York, N. Y., Jul. 1928, pp. 438–534.

Electrical Formulas in Inch Units of Length

J. W. WILLIAMSON
MEMBER AIEE

ELECTRICAL formulas in textbooks and periodicals are usually expressed in metric units of length (centimeters or meters). In many countries, including the United States, dimensions on shop drawings must be reduced to inches. In the past, the inconvenience and possibility of error in converting from metric to inch units have created at least a minor barrier between electrical theory and practical application. In addition, it has appeared that an engineer starting in industry must familiarize himself with a group of electrical equations which, usually being expressed in inch units of length, differed slightly from those learned in college.

Prior to the advent of the rationalized meter-kilogram-second (mks) system (properly referred to as the Giorgi system), it was indeed necessary to insert factors such as 2.54 and 10^{-8} into equations in which distances were in inches and electrical quantities in practical units. When such factors are combined with other factors, such as 2π and $\pi/4$, the resulting proportionality constants may be numerical expressions which have little meaning to the engineers who use them, with the result that these engineers may regard their work as a "crank-turning" process rather than the application of professional knowledge.

The break or divsion between industrial practice and educational training may be minimized if equations derived in metric units can be utilized without change when distances are in inches. It is shown that equations derived in the rationalized form (recommended by the International Electrotechnical Commission) of the mks system apply with distances in inches if appropriate values are used for the permeability μ_v and permittivity ϵ_v of vacuum.[1] New units are required only for mechanical force (the joule per inch, or "jupi") and mass (the "jupima").

The IJS Electrical Unit System

Being based on the inch as a unit of distance, the jupima as a unit of mass, and the second as a unit of time, the system of electrical units introduced in this paper is called the "ijs" system. It might also be called the "ijsa" system since currents are expressed in amperes and other electrical quantities in corresponding practical units.

The ijs system is merely a variation of the rationalized mks system in which distances are expressed in inches rather than meters. Table I gives a comparison between the two systems and also a comparison with a third rationalized unit system in which distances are expressed centimeters.

The Rationalized MKS and IJS Electrical Unit Systems

Such confusion exists about electrical units that a brief discussion or rationalized unit systems and how electrical formulas can be converted from the older unit

---

Paper 56-131, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted August 29, 1955; made available for printing December 23, 1955.

J. W. WILLIAMSON is a consultant electrical engineer, Cleveland, Ohio.

Table I. Three Systems of Rationalized Electrical Units

Characteristic	Name of System Rationalized Giorgi Meter-Kilogram-Second (mks)	 Centimeter-Gramseven-Second	 Inch-Jupima-Second (ijs)
Unit of potential difference	volt	volt	volt
Unit of magnetic flux	weber $=10^8$ maxwells	weber	weber
Unit of magnetic flux density, B	weber/m²	weber/cm²	weber/in.²
Unit of current	ampere	ampere	ampere
Unit of magnetic intensity, H (amperes = ampere-turns)	$\frac{\text{ampere}}{\text{meter}}$	$\frac{\text{ampere}}{\text{centimeter}}$	$\frac{\text{ampere}}{\text{inch}}$
Permeability of vacuum = B/H in vacuum = μ_v	$4\pi\times10^{-7}$ $\frac{\text{webers/m}^2}{\text{amp/m}}$	$4\pi\times10^{-9}$ $\frac{\text{webers/cm}^2}{\text{amp/cm}}$	3.19×10^{-8} $\frac{\text{webers/in.}^2}{\text{amp/in.}}$
Unit of charge to which each electrostatic flux line corresponds	1 coulomb	1 coulomb	1 coulomb
Unit of electric displacement, D	coulomb/m²	coulomb/cm²	coulomb/in.²
Unit of electric intensity, E	volt/m	volt/cm	volt/in.
Permittivity of vacuum = D/E in vacuum = ϵ_v	8.85×10^{-12} $\frac{\text{coulombs/m}^2}{\text{volt/m}}$	8.85×10^{-14} $\frac{\text{coulombs/cm}^2}{\text{volt/cm}}$	2.25×10^{-13} $\frac{\text{coulombs/in.}^2}{\text{volt/in.}}$
Unit of energy	volt Amp Sec = joule $=10^7$ ergs	joule	joule
Unit of force	joule/m = newton $=10^5$ dynes	joule/cm = dyneseven $=10^7$ dynes	joule/in. = jupi = 8.85 pound
Unit of time	second	second	second
Unit of mass	$\frac{\text{newton}}{\text{m/sec}^2}$ $=10^3$ grams = 1 kilogram	$\frac{\text{dyneseven}}{\text{cm/sec}^2}$ $=10^7$ grams = gramseven	$\frac{\text{jupi}}{\text{in./sec}^2}$ $=1.55\times10^6$ grams = 1 jupima

systems (which required that distances be expressed in centimeters) to the rationalized mks system or ijs system seems necessary. Consider the following equations

$$B=\mu_v H \text{ (in vacuum)} \tag{1}$$

$$D=\epsilon_v E \text{ (in vacuum)} \tag{2}$$

If corresponding definitions are adopted for B, H, D, and E, the equations 1 and 2 and all equations derived therefrom, apply in any consistant system of units.

Define a rationalized unit systems as one in which:

1. B is expressed in magnetic flux lines per unit area such that a rate of change of one magnetic flux line per unit time results in one unit of potential difference per turn.
2. H is expressed directly in current turns per unit length.
3. D is expressed in electrostatic flux lines per unit area such that one electrostatic flux line emerges from each unit of positive charge.
4. E is expressed in potential drop per unit length.
5. The unit of energy is equal to the unit of charge multiplied by the unit of potential difference.
6. The unit of mechanical force is equal to the unit of energy divided by the unit of distance.
7. The unit of mass is such that the unit of force, acting on the unit of mass, produces a unit acceleration.

These statements, together with selection of any four basic units (such as units of distance, mass, time, and current), are found completely to define an electrical unit system. All equations in any rationalized electrical unit system (as defined in the foregoing) are identical. Therefore, an equation derived in any rationalized unit system applies in any other rationalized unit system.

According to the definitions of a rationalized system, if potential differences are in volts, magnetic flux lines must be in webers (1 weber $=10^8$ ordinary magnetic flux lines or maxwells), so that a rate of change of one magnetic flux line per second will produce one volt per turn. Therefore, in rationalized mks units, the magnetic flux density B is in webers per square meter. The magnetic field strength H is in ampere-turns per meter; the electric displacement D in electrostatic flux lines per square meter each electrostatic flux line emerging from a positive charge of 1 coulomb; the electric field strength E is in volts per meter; the unit of energy in 1 coulomb × 1 volt = 1 joule; the unit of force is a joule per meter, which is also called a newton; and the unit of mass is such as to require a force of one newton to produce an acceleration of 1 meter per second squared (m/sec²). This unit of mass turns out to be a kilogram.

In the ijs system, potential differences are again in volts, and time is again in seconds. Therefore, magnetic flux lines are in webers. The magnetic flux density B is in webers per square inch. The magnetic field strength H is in ampere-turns per inch; the electric displacement D in electrostatic flux lines per square inch, each electrostatic flux line emerging from a positive charge of one coulomb; the electric field strength E is in volts per inch; the unit of energy is again 1 coulomb × 1 volt = 1 joule; the unit of force is a joule per inch, which is called a "jupi"; and the unit of mass (called a "jupima") is such as to require a force of 1 jupi to produce an acceleration of 1 inch per second squared (in./sec²).

Permeability and Permittivity of Vacuum

The permeability of vacuum is defined as the permeance of a magnetic flux path 1 unit of length × 1 unit of length in cross-sectional area and 1 unit of length long. There is very nearly 0.0254 meter in an inch. It follows that the permeability of vacuum expressed in inch units of length is very closely $0.0254^2\times(1/0.0254)=0.0254$ times as great numerically as the permeability of vacuum expressed in meter units of length. Similarly, the permittivity of vacuum in inch units of length is approximately 0.0254 times as great numercially as the permittivity of vacuum in meter units of length. Hence, starting with values of permeability and permittivity in metric units,

$$\mu_v=4\pi\times10^{-7}\,\frac{\text{webers/meter}^2}{\text{amp (turn)/meter}} \text{ or}$$

$$4\pi\times10^{-7}\,\frac{\text{henries}}{\text{m}} \text{ (with distances in meters)}$$

or

$$\mu_v=3.1919\times10^{-8}\,\frac{\text{webers/inch}^2}{\text{amp (turn)/in.}} \text{ or}$$

$$3.1919\times10^{-8}\,\frac{\text{webers}}{\text{in.}} \text{ (with distances in inches)}$$

and

$$\epsilon_v=8.8549\times10^{-12}\,\frac{\text{coulombs/m}^2}{\text{volt/m}} \text{ or}$$

$$8.8549\times10^{-12}\,\frac{\text{farads}}{\text{m}}$$

(with distances in meters)

or

$$\epsilon_v=2.2491\times10^{-13}\,\frac{\text{coulombs/in.}^2}{\text{volt/in.}} \text{ or}$$

$$2.2491\times10^{-13}\,\frac{\text{farads}}{\text{in.}}$$

(with distances in inches)

Conversion of Electrical Equations from Other Unit Systems

Frequently it may be desired to convert an equation derived in one of the older unit systems to make it apply in rationalized mks or ijs units. Since these older systems were designed to make B numerically equal to H in vacuum (the centimeter-gram-second (cgs) electromagnetic and practical systems), D numerically equal to E in vacuum (the cgs electrostatic system); or both $B=H$ and $D=E$ in vacuum (the Gaussian and Heaviside-Lorentz systems), equations derived therein may be suspected of containing what might be called a "suppressed" μ_v or ϵ_v. Referring back to equations 1 and 2, it should be apparent that the numerical values of μ_v and ϵ_v must change if, for example, the unit of distance is changed. Therefore, μ_v and ϵ_v must be included in equations 1 and 2, and in all equations derived therefrom, for a system of electrical units applicable in any consistent units. The test of whether or not an electrical equation is applicable in any consistent units is to determine whether the dimensions of the left side are the same as those of the right side without assuming either inductance or capacitance to be expressible in centimeters. If the dimensions of the left side differ from those of the right side, it may be assumed that either μ_v or ϵ_v must be added in such a way as to make the equation dimensionally correct in a rationalized sense, in this process of course not changing the numerical value of the right side of the equation in the original system of units. As a simple example, the inductance (henries) of an N-turn coil of length l centimeter (cm) wound about a core of cross-sectional area A centimeter squared (cm²), the length l being much greater than the dimensions of the cross section, is sometimes expressed:

$$L=4\pi\times10^{-9}\,\frac{N^2A}{l}$$

The dimensions of the left side are henries, while the dimensions of the right side are centimeters. However, in reference to Table I, the permeability of vacuum in the practical electrical unit system based on the centimeter as a unit of length is

$$\mu_v=4\pi\times10^{-9}\,\frac{\text{webers/cm}^2}{\text{amp (turn)/cm}}=4\pi\times10^{-9}\,\frac{\text{henries}}{\text{cm}}$$

Hence

$$L=\mu_v\,\frac{N^2A}{l}$$

is correct both numerically and dimensionally and applicable in any rationalized unit system.

As a more complicated example, the "equivalent depth of current penetration" of induction heating is frequently written:

$$d=(1/2\pi)\sqrt{\rho\times10^9/f}$$

d being the depth of current penetration in centimeters, ρ the resistivity of the material in ohm-centimeters, and f the frequency in cycles per second. This equation also is not dimensionally correct in a rationalized sense, but may be replaced by

$$d=\sqrt{\rho/(\pi f\mu_v)}$$

which is dimensionally correct and applicable in any rationalized unit system.

The mks system is not always used in its rationalized form. For example, some well-known books by the staff of the Massachusetts Institute of Technology express the magnetic field strength H in units (praoersteds) such that H is equal to the ampere-turns per meter multiplied by 4π, and the electric displacement D in an electrostatic flux lines per unit area, each of which emerges from a positive charge of $(1/4\pi)$ coulombs. The definitions of magnetic flux density B, electric field strength E, and all other quantities are identical with those in the rationalized mks system. Letting μ_0 stand for B/H and ϵ_0 stand for D/E in unrationalized mks units (all in vacuum); and μ_v and ϵ_v stand for the same ratios in rationalized mks units,

$$\mu_0=(1/4\pi)\mu_v \tag{3}$$

$$\epsilon_0=4\pi\epsilon_v \tag{4}$$

Equations derived in unrationalized mks units apply in rationalized mks or ijs units unless they contain any of the four quantities H, D, μ_0, ϵ_0. In this case the differences in definition and numerical value must be kept in mind in rationalizing.

Units of Force and Mass

In referrence to the characteristics of a rationalized electrical unit system (some of which apply to unrationalized systems also), the unit of force is equal to the unit of energy divided by the unit of distance. Hence, forces come out in joules per inch or newtons in the mks system, or joules per inch or "jupis" in the ijs system.

$$1\text{ jupi}=\frac{1\text{ joule}}{1\text{ in.}}=\frac{10^7\text{ dyne cm}}{2.54\text{ cm}}=(1/2.54)\times10^7\text{ dynes}=8.851\text{ pounds}$$

The unit of mass such as to require a jupi for an acceleration of 1 in./sec² is

$$1\text{ jupima}=\frac{1\text{ jupi}}{1\text{ in./sec}^2}=\frac{(1/2.54)\times10^7\text{ dynes}}{2.54\text{ cm/sec}^2}=(1/2.54)^2\times10^7\text{ grams}=1.5500\times10^6\text{ grams}$$

The approximate mass of an electron is 9.10×10^{-28} grams $=5.87\times10^{-34}$ jupimas.

Conclusions

The system of electrical units proposed (the ijs system) is not intended to replace the rationalized mks system, but to form an auxiliary system in which all equations are identical with those of the rationalized mks system. This permits utilization without change of equations derived in metric units, with distances in inches. The joule per inch or "jupi" as a unit of force and the "jupima" the corresponding unit of mass, may be used by electrical engineers in practical design calculations in much the same way that mechanical engineers express mass in "geepounds" or "slugs." The result in both cases is that equations derived in metric units in theoretical texts apply also in units convenient for every day use.

Some of the units in the ijs system appear inconveniently large. The magnetic flux density B comes out in webers per square inch. The milliweber per square inch would be a convenient unit for industrial design data, since the saturation flux density of iron at room temperature is approximately 1.4 milliwebers per square inch; it is worth remembering that the corresponding rms flux density is approximately 1 milliweber per square inch. The unit of electrostatic flux density D is also very large, but this appears to introduce no problems. In making a magnetic or electrostatic flux plot, one line or tube of force would stand for only a small fraction of a line of B or D as defined in this paper. While the "jupima" = 1,550 kilograms might be considered an inconveniently large unit of mass for a particle as small as an electron, actually it makes little difference whether the mass of the electron be expressed in milligrams or tons, since multiplication by 10 raised to an exponent is required for convenience in either case.

An equation derived in one of the older unit systems may be rewritten in rationalized form. When the equation is rewritten, care should be taken that all quantities involved are defined according to the rules stated in the foregoing for a rationalized system of electrical units. The equation should be checked to determine whether the dimensions of the left side reduce to those of the right side without assuming inductance or capacitance to be expressible in units of length, such as centimeters. If the dimensions of the left and right side do not agree, the permeability μ_v or permittivity ϵ_v of vacuum must be inserted in such a way as to make the dimensions of the left side the same as those of the right side. The manner in which μ_v or ϵ_v is inserted must, of course, be such as not to change the numerical value of the right side of the equation in the original system of units.

While the permeability and permittivity of vacuum may be converted from metric to inch units by following familiar rules, such as would be used for the conversion of electrical resistivity, the approximate numerical values should be memorized. These are

$$\mu_v = 3.19\times10^{-8}\,\frac{\text{webers/in.}^2}{\text{amp (turn)/in.}}$$

$$= 3.19\times10^{-8}\,\frac{\text{henries}}{\text{in.}}$$

$$\epsilon_v = 2.25\times10^{-13}\,\frac{\text{coulombs/in.}^2}{\text{volt/in.}}$$

$$= 2.25\times10^{-13}\,\frac{\text{farads}}{\text{in.}}$$

References

1. Recommendations of IEC Technical Committee 24; Electric and Magnetic Magnitudes and Units, J. J. Smith. *Electrical Engineering*, vol. 74, May 1955, pp. 406–08.

2. Handbook of Engineering Fundamentals (book), O. W. Eshbach. Section 3, John Wiley & Sons, Inc., New York, N. Y., first edition, second edition, 1936, 1952.

A New Technique for the Measurement of Corona Field Strength and Current Density in Electrical Precipitation

P. COOPERMAN
NONMEMBER AIEE

IN THE course of studies of the electrical precipitation of gas-borne dust, it became necessary to determine the electric-field strength in the negative space-charge region of a corona discharge in room air. However, no satisfactory technique for making such measurements was known to the writer or to the persons whom he consulted, nor did a brief survey of the literature offer anything of use. A number of attempts were made to use static probes of types encountered in electrostatics, but these were uniformly unsuccessful. Hence, it became necessary to develop an instrument based on a new principle. This principle makes use of the laws governing the charge picked up by a particle in a unipolar space-charge field as a function of field strength, current density, and time. Essentially, the quantities measured are the particle charge and the time spent by a particle in the field. By the use of the particle-charging laws, the values of the field strength E and of the current density i may be deduced.

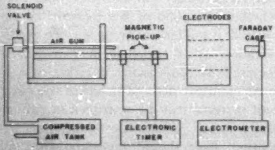


Fig. 1. Schematic diagram of the apparatus

The technique was not pushed to a high level of accuracy because this was not necessary to the use for which it was intended. Nevertheless, it is felt that, in spite of this defect, the possibilities of this method make publication desirable.

Theory

The probe operates by firing precision steel ball-bearings of 3/16-inch diameter from an air-gun, through the discharge under investigation, and into a Faraday cage where the charge picked up is measured by an electrometer. A schematic drawing of this arrangement is shown in Fig. 1. The charge picked up by a steel ball is subject to the laws of particle-charging in unipolar fields[1] which for present purposes may be written, in electrostatic units (esu)

$$V^{-1} = V_s^{-1}(1+v/v_0) \tag{1}$$

$$V_s = 3a^2E/C \tag{2}$$

$$v_0 = \pi Lj/E \tag{3}$$

where

V = voltage in Faraday cage
v = velocity of ball
C = capacitance of Faraday cage + electrometer + cables
L = length of trajectory within discharge
a = radius of ball

Equation 1 indicates that a plot of V^{-1} against v for spheres projected along a given trajectory should be a straight line with intercepts V_s^{-1} and $-v_0$. These quantities determine E and i through equations 2 and 3. A sample plot is shown in Fig. 2.

The foregoing shows that where more than two velocities are used for a given trajectory, deviation of the plotted points from a straight line is an indication of experimental error, or of an incompatibility of the experimental conditions with the hypotheses of particle-charging theory. Thus the new measurement technique to a certain extent has proved self-checking.

Apparatus

The air-gun consists of a barrel 2 feet in length, with an inside diameter of 3/16 inch (+0.001, −0.000 inch), a pressure tank storing air in the range of 0–125 pounds-per-square-inch gauge, and a solenoid control valve. The barrel is

---

Paper 56-113, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 24, 1955; made available for printing December 27, 1955.

P. Cooperman is with Research-Cottrell, Inc., Bound Brook, N. J.

Among the many persons without whose help this work would have been impossible, the author would like to thank H. J. White for his valuable suggestions covering the entire range of the project, J. A. Armstrong for his efforts in developing a suitable air-gun, and J. C. Shepard who built the apparatus used in the experiments, and who amassed the data used.

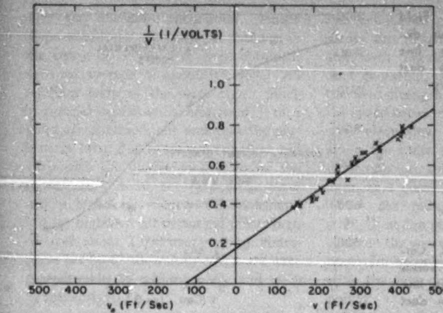


Fig. 2. Sample plot of probe data. Solid line was fitted by eye

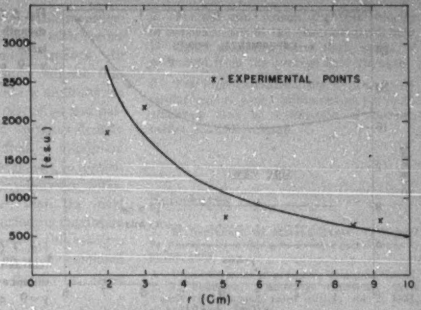


Fig. 4. Current density versus r in coaxial cylinder geometry. Solid line represents values calculated from $j=i/2\pi r$

mounted on a long bar which also contains two magnetic pickups feeding into an electronic timer. The bar, in turn, is mounted in a framework which permits horizontal and vertical adjustments in steps of 1/2 inch. After traversing the discharge, the shots are caught in a Faraday cage. The inner cylinder is of 1-inch diameter and 2-inch length, and is shielded by an outer cylinder of 2-inch diameter and 12-inch length. The back end of the cage is fitted with a coaxial connector from which a cable runs to a Cenco-Dershem electrometer.

It has been found that the usual braided shielding of coaxial cable is not sufficient to prevent pickup of stray ions from the discharge. This is easily overcome by wrapping the cable with aluminum foil. With this precaution, the electrometer holds steady at its zero reading until a shot is fired. After the firing of the shot the electrometer holds the new reading indefinitely.

Experimental Results

The first series of experiments was conducted on a set of coaxial cylinder electrodes consisting of an 8-inch tube and a steel wire of 0.109 inch in diameter. Since the electric field for this geometry has already been investigated both theoretically[1,2] and experimentally,[1] the object of these tests was to compare the measured values of E with those calculated from the equation, in esu

$$E^2=(r_0E_0/r)^2+2i/k \tag{4}$$

where

E_0 = "critical" field strength for starting corona
r = radial distance from center of wire
r_0 = wire radius
i = current per centimeter (cm) length of electrode
k = ion mobility

E_0 was calculated from the semiempirical equation given by Peek.[3] The mobility k was taken to be 660 cm² per statvolt seconds, in accordance with data given by Loeb.[4] Comparison of measured and calculated values of E was the only quantitative check on the results obtained by this probe, since attempts to measure E by other methods were not successful.

Table I. Comparison of Calculated and Measured Values of E and j at Various Voltages for r=9.2 Cm

Corona Voltage, Kv	E Measured	E Calculated	j Measured	j Calculated
40	5.9	5.4	360	250
50	11.1	10.4	750	590
60	14.0	15.1	1,270	1,250

All values are in esu except for the voltage.

Fig. 3 is a graph of E versus r for an applied voltage of 50 kv. The curve represents the values computed from equation 4, while the experimental re-

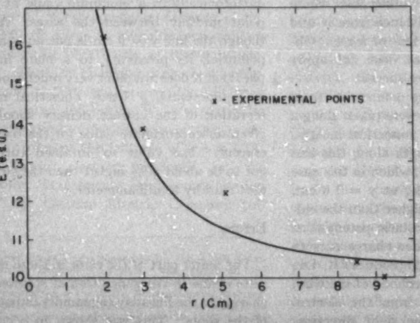


Fig. 3. Field strength versus r in coaxial cylinder geometry. Solid line represents theoretical values

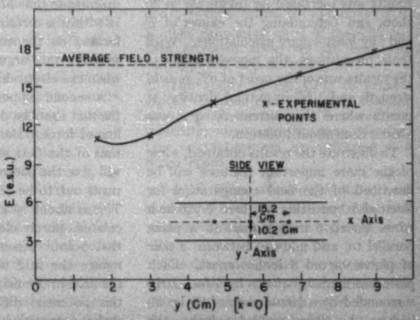


Fig. 5. Field strength versus distance along line x=0 cm in duct geometry

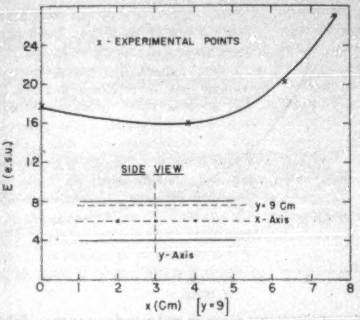
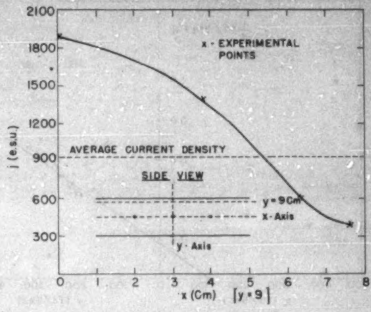


Fig. 6 (left). Field strength versus distance along line y=9 cm in duct geometry

Fig. 7 (right). Current-density versus distance along line y=9 cm in duct geometry



sults are indicated by crosses. It can be seen that the maximum deviation of the experimental from the theoretical values is about 10 per cent (%). Fig. 4 is a similar plot of j versus r and the corresponding maximum deviation is about 30%. As an additional check, data were taken on a line for which r was 9.2 9.2 cm. at applied voltages of 40, 50, and 60 kv; the results are given in Table I. Here again the deviation of the measured from the theoretical values of E and j are of the order of 10% and 30% respectively. Notice also that the measured values of E and j at 50 kv are in fair agreement with the previously measured values of 10.0 and 720 esu respectively.

The next series of measurements was conducted on corona discharge in duct geometry. The geometry produces low current densities as compared to the coaxial cylinders, which in turn results in greater scatter in the data. To improve this condition without redesigning the equipment, 0.020-inch wire was substituted for that of 0.109 inch previously used. In addition, the reliability of the results was increased by taking 20 to 30 shots and determining the values of E and j by least-square calculations. With these precautions, the reproducibility of the results was of the order of 5% in field strength and 10% in current density, at points where the current density was greater than about 1,000 esu.

To illustrate the results obtained, some of the more important features will be described of the field configuration for electrodes consisting of three 0.020-inch wires spaced 6 inches apart in a plane parallel to, and midway between, a pair of plates spaced 8 inches apart. Each plate was made up of a 2- by 1-foot strip surrounded by a guard plate of 29 by 32 inches. At 50-kv applied voltage, the average current density was 930 esu. Fig. 5 is a plot of field strength versus distance from the plane of the wires along the line of force running from the center wire to a plate. Note that the field strength is about 10% above average at the surface of the plate, and that it dips to about 36% below average at 2.2 cm from the wire, and then starts rising again.

A strong argument can be given to support these observations. In the immediate vicinity of a flat plate, the lines of force must very closely approximate straight lines perpendicular to the plate. Consider a small rectangle contained between two lines of force, and two equipotential lines consisting of a small portion of the plate itself and another equipotential. For a negative space charge,

$$E_P - E_Q < 0 \qquad (5)$$

where

E_P = field-strength at the plate
E_Q = field-strength at the opposite side of the rectangle

Hence E_P is greater in absolute magnitude than E_Q. This leads to the statement that the field strength, at a point on the surface of a flat plate, is greater in magnitude than at any other point which is within a certain distance from it and located on the same line of force. Obviously, this argument does not apply when the electrodes are curved.

A second supporting point arises from the fact that the data were taken along a line of force. Hence, numerical integration of the field strength along this line will give the potential, which in this case turns out to be 36.7 kv at $y = 1.6$ cm. This is about 10 kv higher than the calculated, purely electrostatic potential at that point. Since space charge acts to reduce the field which produces it, i.e., the field in the neighborhood of the wire, this potential differs from the electrostatic potential in the right direction. This gives at least qualitative support to the results shown in Fig. 5.

Field strength and current distribution along a line parallel to the plane of the wires but perpendicular to the wires themselves, and at a distance of 9 cm, is shown in Figs. 6 and 7. The field strength, starting from a value of about 18 esu opposite the center wires, declines gradually to 16, and then rises suddenly to about 27 esu at a point midway between two wires. This sudden rise has not yet been explained on a theoretical basis. However, other workers in this laboratory, using an electrical precipitator of similar geometry to collect oil-mist droplets, have noticed heavy sootlike deposits on the plates along lines parallel to and midway between corona wires. A plausible explanation of this phenomenon is that the deposits are due to particles of varnish formed from the oil and collected at these points by an exceptionally high field. This view offers partial confirmation for the measured values of field strength. In contrast to this unusual behavior, the current-density curve follows the expected pattern of being maximum at a point opposite a wire and declining monotonically to a minimum value at a point midway between the wires. Although the line $y = 9$ cm is not an equipotential, its proximity to a plate implies that it does not differ very much from an equipotential. Hence, numerical integration of the current density should given an approximate value for the total current. The value so obtained turns out to be about 30% higher than the current read by a milliammeter.

Errors

The major part of the error in these results is due to the generation of spurious charges in the Faraday cage under impact of the shots. This was shown to occur by firing shots through a sheet of aluminum foil covering the mouth of the cage with the discharge turned off. Under these conditions, electrometer readings in the range of 0–0.25 volt were obtained, with an average of about 0.1 volt. All readings were of the same sign. Since in normal operation, readings of 1 to 4 volts are obtained, one sees that the percentage error due to spurious charges will vary with the charge picked up by the shots. In general, the percentage error will be high for low current densities and low for high current densities. For those of over about 1,000 esu, ~0.3 milliamperes per foot,[2] the values of E and i were reproducible to within ±5% and ±10 % respectively.

Conclusions

The probe method discussed here has been used to measure field strength and current density in electrical precipitators. For cylindrical geometry, the measurements were in fair agreement with theory and with previous measurements. In duct geometry, the results indicate that the field strength reaches a minimum in the space between the discharge and passive electrodes. The current density along the plate goes from a maximum value opposite a discharge wire to a minimum at a point midway between wires. There are indications that the field strength at the plate surface is maximum between the wires; however, this result is uncertain because of the error associated with the low current density at these points.

These results indicate the following with regard to duct precipitators:

1. The charge on collected particles may be higher than would be expected, since these particles must get into the high-field region near the plates.
2. For the same reason, the drift velocity of particles, due to the electrical field, may be higher near the plates than would be estimated from the average field strength.
3. For high current-density types of precipitators, i.e., $j > 0.1$ milliamperes per foot[2], efficiency may be a function of current, rather than of voltage.

References

1. **M. Pauthenier, M. Moreau-Hanot.** *Le Journal de Physique et le Radium*, Paris, France, series 7, 3, 590, 1932.
2. CONDUCTION OF ELECTRICITY THROUGH GASES, J. J. Thomson. Cambridge University Press Cambridge, England, third edition, vol. 2, 1931, p. 552.
3. DIELECTRIC PHENOMENA IN HIGH VOLTAGE ENGINEERING (book), F. W. Peek, Jr. McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1922, p. 55.
4. FUNDAMENTAL PROCESSES OF ELECTRICAL DISCHARGE IN GASES (book), L. B. Loeb. John Wiley & Sons, Inc., New York, N. Y., 1939, p. 55.

A New Five-Digit Plugboard System for Card-Programmed Calculator

S. B. WILLIAMS
ASSOCIATE MEMBER AIEE

MISS N. M. NOONAN
NONMEMBER AIEE

THE International Business Machines Corporation's card-programmed calculator (CPC) is one of the most extensively used automatic digital computers in the world today and has been discussed in several technical articles.[1-4] Scientific calculations, engineering problems, and accounting procedures have all been applied to the CPC with success. This flexibility may be attributed to the fact that all of the computer elements of the CPC are wired to plugable terminals on removable plugboards as shown in Fig. 1. Thus, by changing wires on the plugboard, the internal design of the CPC may be altered and optimized for a particular application. The purpose of this paper is to present the features of a new CPC plugboard system developed by the Power Transformer Engineering Section of the General Electric Company primarily for engineering design calculations. This system may be used for any calculations where slide-rule accuracy is adequate.

History and Scope

In late 1953, when the application of a a digital computer to the transformer design problem was being studied,[5] it became apparent that although simple, 2-winding transformers could be handled adequately by existing CPC plugboard systems, the extension into 3-winding transformers and autotransformers would require more memory capacity. Several CPC users have increased memory capacity by adding more units. This prospect did not appear attractive either from the standpoint of fitting these devices into the basic machine design or from the standpoint of extra rental fees. For these reasons it was decided to explore the possibilities of a new plugboard system. Since slide-rule accuracy is usually adequate for design calculations, and the location of decimal points is more easily predictable for design calculations than for more abstract mathematical calculations, a 5-digit floating-decimal general-purpose CPC plugboard system was studied and developed. The primary difficulties encountered in this project stemmed from the fact that the CPC is basically a 10-digit machine and many of the machine elements are so designed. This has resulted in some minor inconveniences for the machine programmer but they are far outweighed by the advantages.

Since the 5-digit system has been in use for over 1 year and several 10-digit programs have been directly converted to it, an evaluation of its effectiveness for engineering design applications has been made. The following advantages have resulted:

1. The memory capacity of the CPC is effectively doubled, enabling more difficult and involved problems to be processed on it.
2. The calculating speed of the CPC has been increased about 60 per cent with corresponding time savings.
3. The ability of the CPC to perform logical decisions and determine alternate courses of action has been increased substantially; consequently more decision-type calculations can be performed at a higher speed.

All of these advantages have been secured with no increase in rental fees and with more than adequate accuracy of results.

The discussion of any computer system entails a description of the various computer elements and the ways in which these elements work together to accomplish the desired results. Some of the features which are described in this paper

---

Paper 56-143, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 24, 1955; made available for printing November 23, 1955.

S. B. WILLIAMS and MISS N. M. NOONAN are with the General Electric Company, Pittsfield, Mass.

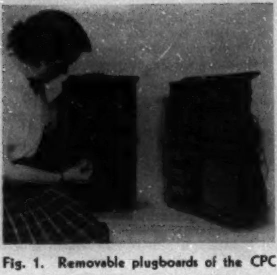


Fig. 1. Removable plugboards of the CPC

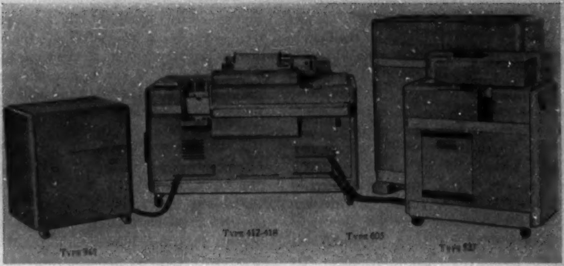


Fig. 3 (above). Machine components of the CPC

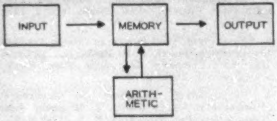


Fig. 2. Basic organization of functions, digital computer

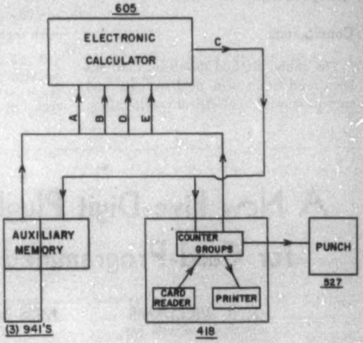


Fig. 4 (right). Schematic organization of the CPC

are common to most CPC plugboard systems and some are peculiar to the 5-digit system under discussion. A knowledge of all of these is necessary, however, for a full understanding of how the 5-digit system operates to perform useful calculations.

Organization

All digital computers from the smallest to the largest so-called giant brains have five main functions: input, memory, arithmetic, output, and control. The first four of these functions are shown schematically in Fig. 2. Control, or the diversion of information from function to function, is implied throughout the entire system. The differences among the various computers available today are in the physical forms that these functions may take, in the capacities of the functions, and in the speed with which information or data can be moved from one to another.

The card-programmed calculator installed in the Power Transformer Engineering Section at Pittsfield, Mass., is shown in Fig. 3 and consists of a type-*418* accounting machine, three type-*941* memory units, a type-*605* electronic calculator and a type-*527* summary punch machine. These units are interconnected in accordance with the organization requirements of a digital computer by appropriate wires in a plugboard system. Fig. 4 shows these interconnections and the location of the functional components which will be described in detail.

INPUT

Input to the CPC is in the form of standard International Business Machine cards, each of which is punched according to a prearranged system. The cards are interpreted by a card reader on the type-*418* accounting machine at a rate of 150 per minute. The time required to read one card (0.4 second) is called a card cycle. There are three different types of cards required to make up a complete CPC program: the control card, the data card, and the instruction card.

The Control Card contains punches which direct or order several mechanical operations of the CPC. For example, such a card might tell the machine to prepare to receive data, print or punch out results, space the printed form, skip to the first line of the next page, etc. A typical control card is shown in Fig. 5.

After the CPC has been told by a control card to prepare to receive data, a data card follows, containing digit punches which tell the CPC what the values of the variables are for the particular calculation being performed. Each datum number which is processed in this system has five digits. The first four digits of each number are the significant digits of that number. The last digit locates the decimal point with the digit 4 selected as the zero base. Therefore, within the range of $(1.000) \times (10)^{-4}$ to $(9.999) \times (10)^{5}$, any number may be converted to an equivalent CPC datum number as follows: Normal number $(n.nnn) \times (10)^{p} = nnnn(p+4)$, the equivalent CPC datum number. This 5-digit number can then be punched onto a data card and readily interpreted by the machine. Negative numbers are indicated by an x double punched into the fifth column of the datum number. The following examples help to illustrate this feature:

Normal Number	Equivalent CPC Datum Number
1	10004
1000	10007
0.001	10001
45.73	45735
−1.23	$1230\overset{x}{4}$

As many as eight 5-digit numbers can be punched into one data card and all of these numbers are transferred to the memory of the machine at one time. A typical data card is shown in Fig. 6.

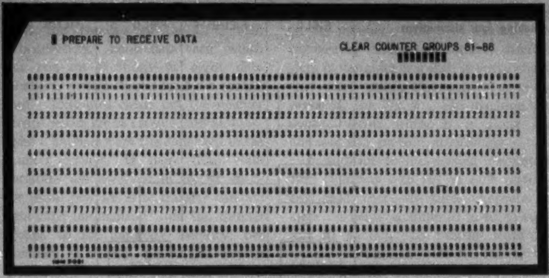


Fig. 5. Typical control card

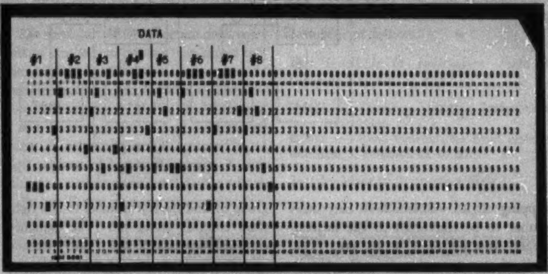


Fig. 6. Typical data card

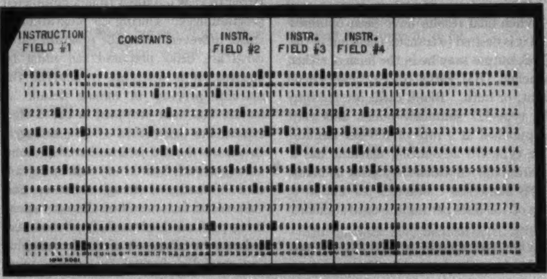


Fig. 7. Typical instruction card

Once the desired data have been transferred to the memory of the CPC, instruction cards follow which tell the machine what to do with the data. These cards contain punches which tell the machine where the data may be found in memory, what arithmetical operations to perform on the data, and where in memory to transfer the result of the operations. To take the machine through the desired calculation, one or several instruction cards would be required. Very difficult calculations sometimes require hundreds or even thousands of such cards. A typical instruction card is shown in Fig. 7.

MEMORY

There are two types of electromechanical memory in the CPC: counter group memory in the type-*418* accounting machine and auxiliary memory in the type-*941* units. The counter groups can store 12 5-digit numbers with their signs. In each *941* unit there is capacity for 16 10-digit numbers with their signs. In the 5-digit system, being described, the 10-digit registers have been apportioned into two 5-digit storage devices, but the two 5-digit numbers in one register must have the same sign. This has not been a serious limitation for the types of calculations performed with this system.

Each of the 5-digit memory positions has been assigned a location number or address. The counter groups are numbered 81–88 and 91–94. The auxiliary memory registers are numbered 11–18, 21–28, 31–38, 41–48, 51–58, 61–68. To differentiate between the left and right halves of the *941* registers, an x is double-punched over the low-order digit of the proper address to indicate the left five digits. In this way, address $2\overset{x}{4}$ refers to the left five digits in register 24 while 24 alone refers to the right five digits. Any of these addresses punched into the proper position on an instruction card will transfer the 5-digit datum number in the memory unit with that address to the electronic calculator. The total memory capacity of the CPC using this system is 108 5-digit numbers. Counters 81–88 have a special significance in the system being discussed. As each data card passes through the machine, direct channels have been provided from the *418* reader to these counters so that the datum number punched in columns 1 to 5 is transferred to counter 81, the datum number in columns 6 to 10 is transferred to counter 82, etc. In this way, eight data numbers may be read into the CPC at one time.

It is interesting to point out that transferring a number from a memory unit to the calculator or to some other location does not destroy that number in the memory unit. It remains there until a different number is directed into that memory unit.

ARITHMETIC

All arithmetical operations for the CPC are performed in the type-*605* electronic calculator. As each instruction card passes through the CPC, four data numbers may be sent into the *605* from either the counter groups or auxiliary memory over four electrical channels provided between these units. These data numbers or factors are called A, B, D, and E. The result of the operations on these factors is called C and it may be sent to the counter groups or auxiliary memory or used as any one of the four factors for the next instruction card. Mathematically, this relationship may be expressed as

$$[(A \text{ op } B)\text{op } D]\text{op } E = C$$

where op may be one of the basic arith-

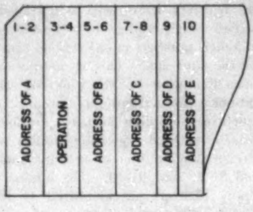


Fig. 8. Instruction field

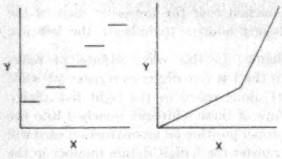


Fig. 9. Typical design functions

metical operations, +, −, ×, or ÷. The 5-digit system, being discussed, makes possible 60 different combinations of arithmetical operations on the four factors during each instruction card cycle.

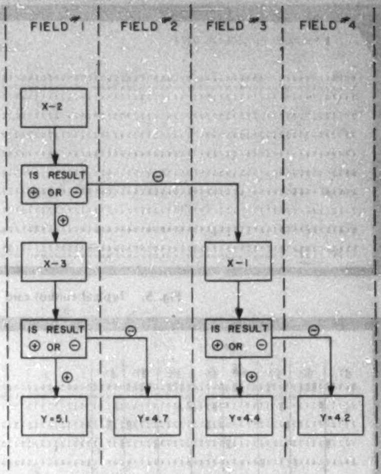
In addition to the arithmetical operations, one of several transcendental functions may be called for on an instruction card. These are:

$\sin A = C$
$\cos A = C$
$e^A = C$
$1 - e^A = C$
$\ln A = C$

Since the *605* calculates these functions by expanding the appropriate power series and many terms of these series are sometimes necessary to obtain the desired accuracy, calculation of transcendental functions requires more than one card cycle. When this is the case, the *418* card reader will stop feeding cards until the calculation is completed. A program deck containing many transcendental operations will run substantially slower than one containing only arithmetical operations.

Two other very useful operations which can be performed in the *605* are dropping digits and rounding. The machine can be instructed to drop all digits to the right of a certain decimal location. This is particularly useful in engineering design applications where many variables are discrete in size.[5] The dropping-digits operation allows any raw number to be converted to the next lower or higher integer, multiple of 2, 4, 10, etc. The machine can also be instructed to round or half adjust any number before dropping digits. This allows any raw number to be converted to the nearest integer, multiple of 2, 4, 10, etc. The result of any operation *C* in the *605* is always in floating decimal form.

Fig. 10. Form of program having four alternatives



Output

When final results have been obtained and it is desired to transfer them out of the CPC, output may be in the form of either punched IBM cards or a printed document, or both. Blank cards are fed into the type-*527* summary punch and may be punched at a rate of 100 per minute. The type-*418* accounting machine has a printing device which may print as many as 88 numerical characters per line at a rate of 150 lines per minute onto standard forms.

When the counter groups (81–88) are loaded from the data cards, the same data numbers are printed onto the output sheet. This gives the operator or interested engineering specialist an opportunity to check the accuracy of input data. It also provides a check between input data and output data for later reference.

Just as the eight counter groups may be loaded simultaneously from a data card, the same counter groups may be unloaded simultaneously to the *527* punches or to the *418* type-bars or to both. By appropriate punches on a control card, the desired information is made available to the CPC operator in punched or printed form in the same positions as the numbers on a data card, i.e., the first number to the left comes from counter group 81 and the last number to the right from counter group 88. The punched card will look exactly like a data card and may be used as such for future calculations if desired.

All of the output described has referred to that obtained from a completed program deck. During the programming stage, however, when the program cards cards are being prepared and must be checked, it is desirable to analyze the machine program in minute detail. A feature has been provided in the 5-digit system for doing this. When a toggle switch on the *418* is transferred, a complete printed listing of each program card is made, showing addresses and values of all factors and results, operations being performed, etc. This printed record may be taken away from the machine and carefully checked for programming errors.

Card Program Deck

The sequence of instructions to a digital computer directing it to perform a complete calculation is called a "program." With the CPC, this program consists of control, data, and instruction cards combined into a set of cards called a program deck.

The size of a program deck varies with the length and complexity of the calculation being performed; however, the form of each program deck is usually the same. Since no calculation can be performed until numbers are available on which to operate, the first part of a program deck loads the memory of the CPC with data numbers. Each data card, which can contain as many as eight data numbers, must be preceded by a control card. This card clears out the counter groups which are to receive data and provides a channel from the data card to the counter groups. If more than eight variables are involved in the problem, the usual practice is to transfer the first data-card information from the counter groups to the auxiliary memory, two numbers at a time, reload the counter groups with a second data card, transfer this information to auxiliary memory, etc., until values for all the variables are located in CPC memory.

The next part of the program deck consists of instruction cards which direct the machine through the calculations required. Each instruction card contains an instruction field as shown in Fig. 8. Columns 1 and 2 contain the 2-digit address of factor *A*, columns 5 and 6 contain the 2-digit address of factor *B*, column 9 contains the double-punched 2-digit address of factor *D* and column 10 contains the double-punched 2-digit address of factor *E*. Columns 3 and 4 contain the operation code which tells the machine what to do with the four factors. Columns 7 and 8 contain the 2-digit address of the result *C*. Up to three arithmetical operations may be performed on each instruction card as explained before. Two other features of the 5-digit system may be discussed here.

CONSTANTS

Columns 11 to 30 of each instruction card are reserved for constant values of the factors *A*, *B*, *D*, and *E*. A special address code tells the machine that the desired number is on the card rather than in memory. For example, if π is required as factor *A* on a particular instruction card, the number 31424 would be punched into columns 11 to 15; if 2 is required as factor *B* on an instruction card, the number 20004 would be punched into columns 16 to 20, etc. Thus, constants do not occupy valuable memory space.

ALTERNATE FIELDS

Columns 31 to 60 of each instruction card are reserved for alternate instruction fields. One of the most powerful features of the digital computer is its ability to recognize the magnitude of a number generated in a calculation and proceed on different courses of action depending on the number's size. For example, in engineering design calculations, discrete steps in variables or a shift in curve shape may occur within the range of a given problem; see Fig. 9. The machine can readily decide where it is along the abscissa and proceed accordingly.

In addition to the normal instruction field (columns 1 to 10) on each instruction card, there are three alternate instruction fields (columns 31 to 40, 41 to 50 and 51 to 60). To transfer from the normal instruction field to one of the alternate fields, an arithmetical operation is performed and the machine is instructed to sense the sign of the result. There are three different tests that may be made, as follows.

Test 1—Made in Instruction Field 1, Columns 1 to 10

Result +, Continue reading instructions from instruction field 1.
Result −, Read instructions from instruction field 3, columns 41 to 50.

Test 2—Made in Instruction Field 1

Result +, Continue reading instructions from instruction field 1.
Result −, Read instructions from instruction field 2, columns 31 to 40.

Tests 3—Made in Instruction Field 3

Result +, Continue reading instructions from instruction field 3.
Result −, Read instructions from instruction field 4, columns 51 to 60.

For example, if an operation of the following form must be programmed:

If $0\leq x<1$, then $y=4.2$
If $1\leq x<2$, then $y=4.4$
If $2\leq x<3$, then $y=4.7$
If $3\leq x$, then $y=5.1$

the CPC program would be of the form shown in Fig. 10; the machine interprets 0 as +.

The machine may be instructed at any time to return to instruction field 1 and continue the normal calculation.

With the use of all the features just described, the desired calculations may be performed. When results have been obtained and stored in memory, it is necessary to get them out of the machine so they may be read and interpreted by the machine operator or interested engineering specialist. The last part of the program deck consists of instruction cards which transfer the results to the counter groups 81 to 88, and a control card which directs the machine to print or print and punch these result numbers out as explained before. If more than eight results are desired, this procedure may be repeated until all results are out of the machine.

Once a program deck has been prepared and thoroughly checked, it is necessary only to punch new data cards for a new run of this same type of calculation. The control and instruction cards remain unchanged. To pick out data cards readily in a program deck, it has been found desirable to use white cards with the upper right corner cut for data cards, and colored cards with the upper left corner cut for control and instruction cards.

There are over 500 different program decks on file at the Pittsfield CPC installation at present. These decks are used for all types of engineering design and developmental and statistical calculations. Practically all of the design decks use the 5-digit general purpose plugboard system described in this paper.

Conclusions

1. The card-programmed calculator is a plugboard-type of digital computer which may be optimized for a particular kind of application.

2. A 5-digit general-purpose floating-decimal plugboard system has been successfully developed.

3. This system may be used for any calculations where slide-rule accuracy is sufficient and it has been found to be excellent for engineering design calculations.

4. The system is easy to program for and easy to check out.

5. The capacity of the CPC has been increased substantially by the use of this system. The memory has been effectively doubled and the computing speed increased by 60 per cent with no increase in machine rental fees.

References

1. The IBM Card-Programmed Electronic Calculator. *Proceedings*, Seminar on Scientific Computation, International Business Machines Corporation, New York, N. Y., Nov. 1949, pp. 37–41.

2. Review of Electronic Digital Computers. *AIEE Special Publication S-44*, "The IBM Card-Programmed Electronic Calculator," John W. Sheldon, Liston Tatum. Feb. 1952, pp. 30–36.

3. Save Time, Money With Computers, A. Opler. *Chemical Engineering*, New York, N. Y., Oct. 1954, pp. 197–208.

4. Solving Problems With the Semac II, D. D. McCracken. *General Electric Review*, Schenectady, N. Y., July 1954, pp. 50–52.

5. Application of Digital Computers to Transformer Design, S. B. Williams, P. A. Abetti, E. F. Magnusson. *AIEE Transactions*, vol. 75, pt. III, 1956 (*Paper No. 56-80*).

Note on the Approximation Problem

NORMAN BALABANIAN
NONMEMBER AIEE

NETWORK synthesis is the aggregate of concepts and procedures utilized in obtaining an electric network whose performance approximates an arbitrarily prescribed performance. The performance specified may be the desired time response to a given excitation, such as a step or impulse function. More usually, it is the magnitude of the steady-state transfer function over all or part of the frequency axis; or it may be the angle of this function specified over all or part of the frequency axis, or it may be some combination of these over parts of the frequency axis. This paper is concerned only with a magnitude function prescribed over part or all of the frequency axis. The network functions of interest are output over input functions, with any combination of voltage or current as output and input. All of these are functions having similar properties. To include them all in the discussion the transfer function F will be defined as a function specifying the output of a network divided by the input.

In general, the network performance only approximates the specified performance because an arbitrarily specified function may not be exactly realizable by an electric network. Thus, the work involved in network synthesis may be divided into two parts:

1. Obtaining a function which approximates the prescribed function to within prescribed tolerances while at the same time being realizable as a network transfer function.
2. Obtaining a network that realizes the function found in the first part.

This paper is concerned with the first part of the synthesis problem.

A prescribed function can be approximated in one of several ways. Three useful types of approximation are:

1. Least mean square approximation. Here, the approximating function is such that the average value of the square of the error between the two functions is a minimum.
2. Taylor approximation. In this type the Taylor series of the prescribed and approximating functions are equal term by term up to a designated power. This yields a so-called maximally flat approximation.
3. Tchebyscheff, or equal ripple, approximation. In this type the approximating function oscillates, or ripples, about the prescribed function in such a way that the maximum deviations between the two functions are all equal.

In this paper a combination of the first two types will be considered.

Some Properties of Transfer Functions

The transfer function of a linear, lumped network is a rational function of the complex frequency, $s=\sigma+j\omega$, given by

$$F(s)=\epsilon^{\theta_e(s)+\theta_0(s)}$$
$$=K\frac{(s-s_1)(s-s_2)\ldots(s-s_q)}{(s-s_1')(s-s_2')\ldots(s-s_m')} \quad (1)$$

where $\theta_e(s)$ is an even function and $\theta_0(s)$ is an odd function of s. For a realizable transfer function the poles must lie in the left half-plane. The magnitude of $F(j\omega)$ can be obtained from the function

$$G(s)=F(s)F(-s)$$
$$=K^2\frac{(s^2-s_1^2)\ldots(s^2-s_q^2)}{(s^2-s_1'^2)\ldots(s^2-s_m'^2)} \quad (2)$$

by noting that

$$|F(j\omega)|^2=G(j\omega) \quad (3)$$

or

$$\theta_e(j\omega)=\frac{1}{2}\ln G(j\omega) \quad (4)$$

When $s=j\omega$, θ_e becomes the usual gain function.

Some of the properties of $G(s)$ are listed here.[1]

1. $G(s)$ is a rational function of s^2, real when s^2 is real, and positive when s^2 is negative.
2. Zeros and poles of $G(s)$ meet the following conditions:

   Complex zeros or poles occur in sets of four having quadrantal symmetry.

   Real zeros occur in pairs symmetrical about the origin.

   Imaginary zeros or poles occur in conjugate pairs of even multiplicity. This applies to zeros or poles at zero and infinity also.

If the zeros and poles of $G(s)$ are known, those of $F(s)$ can be found. Since $F(s)$ is analytic in the right half s-plane, the left half-plane poles of $G(s)$ are assigned to $F(s)$. If the network is designated to be of minimum-phase type, the zeros of $F(s)$ must lie in the left half-plane also. Thus, only one possible assignment of the zeros of $G(s)$ to $F(s)$ can be made. If the network is not so designated, there is a flexibility in the assignment to $F(s)$ from among the zeros of $G(s)$.

Approximation Procedure

The problem under consideration is the following: Given an arbitrary function of frequency, $\overline{G}(j\omega)$, prescribed over a band of frequencies not including $\omega=0$, which will be called the band-pass case, and which is to be the squared magnitude of a network transfer function, to find the poles and zeros of a realizable $G(s)$ so that $G(j\omega)$ approximates the prescribed function in some manner. Also of interest will be the case in which $\overline{G}(j\omega)$ is prescribed over the entire frequency axis. The function may be given either analytically or graphically.

The solution of this problem is obtained by a procedure which is similar to the work of Darlington.[2] Darlington deals essentially with a frequency band centered about $\omega=0$ (the low-pass case). His method handles prescribed gain or angle, or simultaneous gain and angle functions.

Darlington's procedure applied to a low-pass gain function (logarithm of the magnitude) is essentially as follows: The transformation

$$s=\frac{1}{2}\left(z-\frac{1}{z}\right);\ z=\epsilon^{u+jv} \quad (5)$$

which becomes $\omega=\sin v$ for $s=j\omega$, transforms the frequency band of interest to the unit circle in the z plane. This transformation is illustrated in Fig. 1. The prescribed gain function is expanded in a series of Tchebyscheff polynomials in ω. The series of Tchebyscheff polynomials in ω is also a Fourier series in v which can be written as a power series in z with positive and negative powers. Furthermore, equation 5 transforms a gain function in s to the sum of two functions in z, one of which has no singularities inside the unit circle, while the other has none outside. In terms of $G(s)$, use of equation 5 gives

$$G(s)=g(z)g\left(-\frac{1}{z}\right) \quad (6)$$

and

Paper 56-128, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted July 5, 1955; made available for printing November 22, 1955.

NORMAN BALABANIAN is with Syracuse University, Syracuse, N. Y.

The author wishes to acknowledge the help of Prof. W. R. LePage of Syracuse University, through exhaustive discussions with whom many of the ideas contained in this paper were developed.

Vol.
Pt. 1
1956

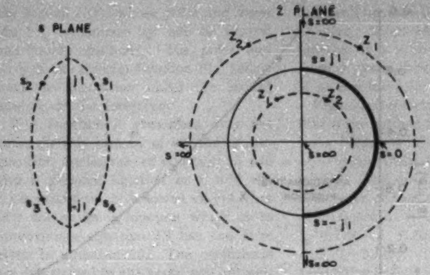


Fig. 1. Properties of the transformation $s = 1/2\,(z - 1/z)$

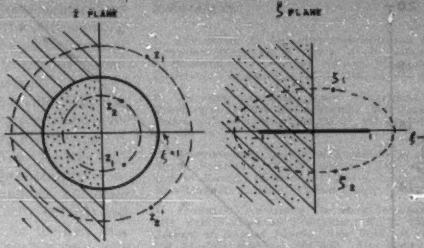


Fig. 3. Properties of the transformation $\zeta = 1/2(z + 1/z)$

$$\ln G(s) = \ln g(z) + \ln g\left(-\frac{1}{z}\right)$$

where $g(z)$ is a rational function in z with no singular points inside the unit circle. Thus, the function in z which has no singularities inside the unit circle can be approximated by a finite sum of the series in positive powers in z. This determines the poles and zeros of the gain function in the z plane and those in the complex frequency plane are obtained from equation 5.

For band-pass functions possessing geometric symmetry the same procedure can be applied by first using the band-pass to low-pass transformation

$$p = \frac{1}{2}\left(s + \frac{1}{s}\right) \qquad (7)$$

where p and s are the low-pass and band-pass variables respectively. However, if a band-pass function does not have geometric symmetry, use of equation 7 will result in a double-valued function in the low-pass plane and the procedure cannot be used.

Recently Grinich[3] has reported an interesting method extending Darlington's procedure to band-pass gain and angle functions. Kuh[4] has also dealt with the problem with the help of the potential analogy. However, both of these procedures involve elliptic functions. This paper gives a rather different method for approximating arbitrary band-pass gain or magnitude functions.

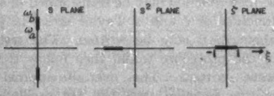


Fig. 2. The pass bands in the s, s^2, and ζ planes using the transformation $\zeta = As^2 + B$

The method is based on some simple transformations. An arbitrary band-pass function in ω can be transformed to a single valued low-pass function by the transformation

$$\zeta = As^2 + B;\ \zeta = \xi + j\eta \qquad (8)$$

where

$$A = \frac{2}{\omega_b^2 - \omega_a^2} \qquad B = \frac{\omega_b^2 + \omega_a^2}{\omega_b^2 - \omega_a^2}$$

This is illustrated in Fig. 2. The upper and lower edges of the band are ω_b and ω_a respectively. The frequency band of interest is transformed to the low-pass interval on the real axis of the ζ plane, $-1 \leq \xi \leq 1$. This interval in the ζ plane can be transformed to the unit circle in the z plane by the transformation

$$\zeta = \frac{1}{2}\left(z + \frac{1}{z}\right) \qquad (9)$$

which for real frequencies becomes

$$\zeta = \cos v \qquad (10)$$

This transformation is illustrated in Fig. 3. Note the similarities with Fig. 1. Thus, the frequency band of interest is transformed to the unit circle in the z plane.

It will be seen how equation 8 transforms a rational function of s^2 as given in equation 2 to a function of ζ. Each factor of the form $(s^2 - s_i)$ becomes $(\omega_b^2 - \omega_a^2)/2(\zeta - \zeta_i)$, so that the rational function $G(s)$ becomes

$$G''(\zeta) = K^2\left(\frac{\omega_b^2 - \omega_a^2}{2}\right)^{q-m} \frac{(\zeta - \zeta_1)\ldots(\zeta - \zeta_q)}{(\zeta - \zeta_1')\ldots(\zeta - \zeta_m')}$$

$$= \frac{1 + a_1\zeta + a_2\zeta^2 + \ldots + a_q\zeta^q}{b_0 + b_1\zeta + \ldots + b_m\zeta^m} = \frac{P_1(\zeta)}{P_2(\zeta)} \qquad (11)$$

A similar procedure can be used if the magnitude or gain function is prescribed over the entire $j\omega$ axis. The $j\omega$ axis of the s-plane is transformed to a low-pass interval by the transformation

$$s^2 = \frac{\zeta - 1}{\zeta + 1} \qquad (12)$$

The interval of interest in the ζ-plane is again the real axis from -1 to $+1$. Use of equation 9 will again transform this interval to the unit circle in the z plane. The relationship between the s and z planes can be obtained by combining equations 9 and 12. It is

$$s = \frac{z - 1}{z + 1} \qquad (13)$$

which for imaginary values of s becomes

$$\omega = \tan v/2 \qquad (14)$$

This is a well-known transformation; see Fig. 4. Perhaps its most familiar application is in the Smith chart.

Equation 12 will also transform the rational function of equation 2 into a rational function in ζ but the form will be slightly different from equation 11. Since the point at infinity in the s plane is transformed to the point $\zeta = -1$, each zero of $F(s)$ at infinity will lead to a zero of $G''(\zeta)$ at $\zeta = -1$. Thus, the rational function of equation 2 will be transformed to

$$G''(\zeta) = (\zeta + 1)^{m-q} \frac{1 + a_1\zeta + \ldots + a_q\zeta^q}{b_0 + b_1\zeta + \ldots + b_m\zeta^m} \qquad (15)$$

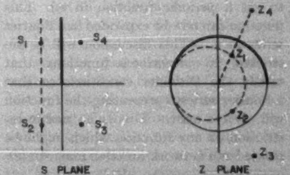


Fig. 4. Properties of the transformation $s = (z - 1)/z + 1$

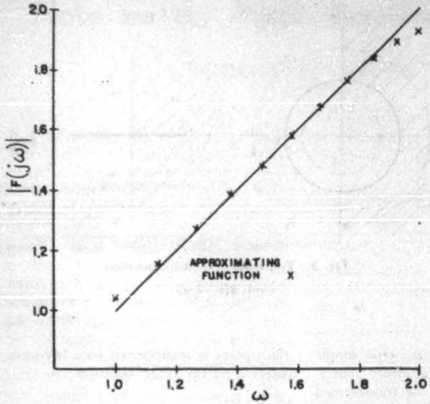


Fig. 5. Approximation of the function $|F(j\omega)|=\omega$ for $1\leq\omega\leq 2$

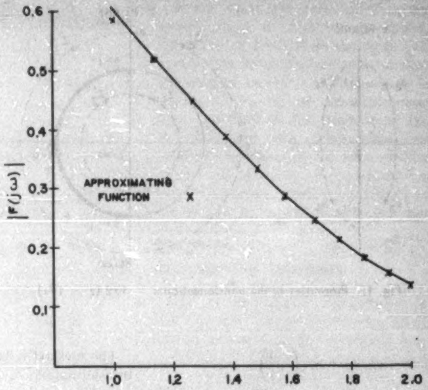


Fig. 6. Approximation of the function $|F(j\omega)|=\exp(-\omega^{2}/_{2})$ for $1\leq\omega\leq 2$

With these transformations Darlington's procedure can now be applied as described previously. However, in some cases it may not be convenient to deal with the gain (logarithm of magnitude) function. For example, it may be prescribed that the magnitude fall to zero or to a very small value at the edges of the band. In such a case, the gain will go to negative infinity or to a large negative value at the edges of the band. The Fourier series representing the gain function will then converge very slowly, if it converges at all, and a large number of terms will be needed to obtain an adequate approximation. This disadvantage can be overcome if the magnitude function is used. The problem then is that of a prescribed magnitude function either over a finite band of frequencies or over the entire $j\omega$ axis.

In this case, the approximation procedure would be as follows. If a magnitude function $\bar{G}(j\omega)$ is prescribed over a band of frequencies, use equations 8 and 10 to transform this to a periodic function of v, $\bar{G}'(v)=\bar{G}(j\omega)$. (If $\bar{G}(j\omega)$ is prescribed for all ω, use equation 14 to get a periodic function in v.) This function can now be expanded in a Fourier series in v. (This discussion will be restricted to prescribed functions that satisfy the Dirichlet conditions so that a Fourier series representing the function will always exist. This is no undue restriction on any function which might be wanted as a network function magnitude).

$$\bar{G}'(v)=\sum_{k=0}^{\infty}\bar{C}_k\cos kv \qquad (16)$$

An approximation to the prescribed function is made by retaining a finite number of terms of this series, the actual number of terms depending on the prescribed tolerance. This finite trigonometric sum is then rewritten as a trigonometric polynomial in $\cos v$ by use of the identity

$$\cos kv=2\cos(k-1)v\cos v-\cos(k-2)v \qquad (17)$$

Finally, use of equation 10 changes this to a polynomial in ζ, which may be written

$$\bar{G}''(\xi)=\sum_{k=0}^{n}d_k\xi^k \qquad (18)$$

This is an approximation to the prescribed function, suitably transformed. The objective is to have this polynomial approximately equal to the rational function of equation 11. The approximation could take one of several forms. The simplest procedure would be to make the coefficients of the Taylor series of $G''(\zeta)$ about $\zeta=0$ coincide with the coefficients in equation 18 up to the nth power for real values of ζ. This can be accomplished by writing

$$\frac{P_1(\xi)}{P_2(\xi)}=\frac{1+a_1\xi+\ldots+a_q\xi^q}{b_0+b_1\xi+\ldots+b_m\xi^m}\doteq\sum_{k=0}^{n}d_k\xi^k \qquad (19A)$$

or

$$P_1(\xi)\doteq P_2(\xi)\sum_{k=0}^{n}d_k\xi^k \qquad (19B)$$

To determine the coefficients, first multiply the right-hand side out. The left-hand side is of degree q. Hence, coefficients of ξ of degree $(q+1)$ to $(q+m)$ on the right-hand side are equated to zero. This gives a set of m simultaneous linear equations in the b coefficients. After this set is solved and the b coefficients determined, the a's are determined by equating coefficients of like powers of ξ from zero to q on both sides of equation 19. To solve for all the coefficients uniquely, n must equal $m+q$. The coefficients are easiest to determine when all the zeros of the rational function $G''(\zeta)$ are at infinity; then $P_1(\zeta)=1$. In this case there is no need to solve simultaneous equations.

There are two approximations involved in the over-all procedure. One is a least mean square type, when a finite number of terms is retained in the Fourier series. The second is a Taylor type, when the coefficients of the rational function are determined. The combination seems to give an approximation which is almost Tchebyscheff, and this has been called quasi-Tchebyscheff or nearly-equal-ripple.

The procedure can be modified to obtain a different type of approximation. Instead of transforming the prescribed function of ω to a function of v and then obtaining the Fourier series, the prescribed function may be transformed to a function of ξ and a Taylor series in ξ about $\xi=0$ may be written. The rest of the procedure would then follow the same pattern. The over-all approximating function then will not oscillate about the prescribed function.

After the coefficients of $G''(\zeta)$ are determined, use of equation 8 will give the desired $G(s)$, whose poles and zeros can be determined by one of several root-finding methods. The poles and zeros of the transfer function $F(s)$ follow from the discussion under the second heading in the foregoing.

If a function is prescribed over the entire $j\omega$ axis instead of a band of frequencies, application of equation 14 will give a periodic function in v, and the same procedure is followed except that now the rational function which is to approximate equation 18 has the form given in equation 15. The coefficients are determined in the same way but now there will be more simultaneous equations to solve. Even when the corresponding function in s is to have all its zeros at infinity, we will need to solve a system of m simultaneous equations.

Figs. 5 and 6 show the results of application of the method to the prescribed functions $|F(j\omega)| = |\omega|$ and $|F(j\omega)| = \exp(-\omega^2/2)$ respectively, both in the range $1 \leq |\omega| \leq 2$. The zeros of both transfer functions were chosen at infinity. The poles of $F(s)$ are at $(-0.34055 \pm j2.1502)$ and $(-0.77816 \pm j1.5272)$ for the first case and (-2.147), $(-0.561 \pm j0.888)$ and $(-0.801 \pm j1.702)$ for the second.

References

1. The Role of Analytic Continuation in Network Synthesis, S. Seely, W. R. LePage, N. Balabanian. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 9, 1953, pp. 604–09.

2. Network Synthesis Using Tchebyscheff Polynomial Series, S. Darlington. *Bell System Technical Journal*, New York, N. Y., vol. 31, 1952, pp. 613–65.

3. Approximating Band-Pass Attenuation and Phase Functions, V. H. Grinich. *IRE Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 2, "Circuit Theory," 1954.

4. Potential Analogous Network Synthesis for Arbitrary Loss Functions, E. S. Kuh. *Journal of Applied Physics*, New York, N. Y., vol. 24, July 1953, pp. 897–902. Also *Technical Report No. 44*, Electronics Research Laboratory, Stanford University, Stanford, Calif., 1952.

Conversion of Studio Lighting from Black and White to Color Television

P. W. WYGANT
NONMEMBER AIEE

THIS is the story of the conversion from black and white to color television (TV) lighting at Station WBAP-TV in Fort Worth, Tex. Some problems of such a conversion will be unique at each different station. For example: the problems of integrating a new lighting system with the old installation, the type of programming anticipated, the personnel available and, of course, the ever-varying budget. This paper presents the needs of the station and describes how they were met. First, however, the lighting facilities on hand before the addition of color equipment will be outlined.

Original Equipment

The studio to be converted was constructed in the summer of 1948; see Fig. 1. The studio is 82 feet long, 45 feet wide, and 28 feet high. A pipe grid is mounted 20 feet above the floor and is constructed of $1^1/_4$-inch galvanized pipe arranged in 7-foot squares suspended from the ceiling. A 3-foot catwalk runs around the two sides and the end of the studio, 14 feet above the floor.

The lighting fixtures and control equipment in use just before the conversion to color were essentially those purchased and installed in 1948. There were 18 Television Associates banks of incandescent lamps: see Fig. 2. Each fixture contained 12 lamps of which six were type-*PAR-38* floods, and six were type-*R-40* floods. Separate circuits were used to supply power to the two types of flood lamps. The *PAR*-type lamps were used for modeling, while the *R-40* type of lamp was used for fill light. Each fixture was designed to permit adjustment in azimuth and elevation. A 360-degree turn plus tilt adjustments were made by means of cords running to a central location. Similar turn-and-tilt adjustments were provided for 15 fluorescent fixtures, which each contained six 64-inch slimline tubes. Back light was provided by fresnel spots which were mounted on the catwalk rail. Ten 2-kw spots and 10 750-watt spots were in use.

The total wattage from all fixtures amounted to 62 kw, or approximately 18.5 watts per square foot. In 1948 and 1949 it was the practice of the station to provide 200 foot-candles for use with the 5655-image orthicon tubes. With the fixtures then available it was not possible to provide 200 foot-candles around the four walls of the studio. However, by means of the turn-and-tilt controls, it was possible to concentrate the light on any two or three areas in use at any one time. Later, with the development of the more sensitive 5820-image orthicon pickup tube, only 100 foot-candle studio lighting was required. A 100-foot-candle illumination could be provided uniformly for any area in the studio without turning fixtures.

Original Lighting Switchboard

The switchboard was designed to handle 140 separate circuits. One hundred forty silent-type mercury switches were used to remote-control a similar number of relay contactors. Each switch was rated at 10 amperes while the load contactors handled the load current of 20 amperes. One relay load contactor was used for each load receptacle. Each load circuit was provided with a 20-ampere circuit breaker. The load contactors were fed from two 200-ampere main breakers in the power feed.

With the advent of the 5820-type image orthicon tube there was a lessening of the infrared problem noticeable with the earlier types of pickup tubes. At this stage a dimmer board of 12-kw capacity was constructed and placed into operation. It was used only by those desiring very special lighting effects since its operation necessitated the running of extension cables along the grid from the dimmer to the fixture. Therefore, a great deal of ladder work was required for dimmer effects. In fact, the main objection to the control switchboard by the operating personnel was its position on the catwalk 14 feet above the studio floor. The load contactors were subject to sticking, thus leaving the circuit continually hot until someone filed or replaced the contact points. Fig. 3 shows the location of the original switchboard on the catwalk.

If the system described in the fore-

Paper 56-76, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 17, 1955; made available for printing November 30, 1955.

P. W. Wygant is with Station WBAP, Fort Worth, Tex.

Fig. 1. WBAP-TV studio showing original lighting equipment

going appears cumbersome and inflexible (and it was) note that it was designed in 1947 when there was very little information available on the subject. Briefly, then, this was the situation early in 1954 before the station started adding lighting equipment in preparation for live color originations.

Air Conditioning

As is well known, the summers of Texas are not mild. One of the first problems at Station WBAP-TV was the disposal of the additional heat to be generated by the increased lighting load. A heat exchanger was built on the roof of an adjoining building with connecting ducts feeding into the end of the studio. Two 100-ton compressor units were already in use. One 60-ton compressor unit was added to the existing chilled water lines which were extended to the new heat exhaust. Additional ducts were added above the grid in the studio. The new studio air-conditioning capacity was almost doubled over that of the existing installation.

Primary Power

The substation in use in 1954 belonged to the Texas Electric Service Company and had a 450-kva capacity. The three transformers in use were purchased from the service company. Three similar 150-kva transformers were added to the transformer bank to give a total of 900-kva installed capacity. The total load was not expected to be doubled, but it was felt that the reserve power capacity would suffice for possible future needs. When the new transformers were installed, the complete substation was purchased from the service company. It appeared to be an economical move to purchase the substation since the saving in power costs would pay for the transformers within approximately 6 years.

In the basement power vault a new 150-kva transformer was added to the secondary bus. Fortunately there was space to fit the new transformer in the same room with two similar 150-kva transformers already in use. The new transformer feeds 120 to 208-volt 3-phase 4-wire service to a 400-ampere main circuit breaker. The second transformer feeds two 200-ampere breakers, one of which is the main for lighting in another TV studio.

The Texas Electric Service Company furnishes primary power on its lines to the property line where it is metered at a common point. Then by WBAP lines primary power is sent to the transmitter building for air conditioning and transmitter power. From the substation at the main building, service is provided to the air-conditioning compressors, heat exchangers, motors, and equipment requiring 440-volt service. The secondary bus feeds a 50-kva transformer serving the building and film-laboratory needs; a 100-kva transformer for radio studio equipment; a 150-kva transformer for regulated power to the TV equipment; a 150-kva transformer serving a 200-ampere main circuit breaker to studio No. 2 lighting; and a 200-ampere main breaker to studio No. 1 dimmer circuit. There are 12 50-ampere subbreakers on this circuit. The last 150-kva transformer on the secondary bus feeds the 120- to 208-volt 3-phase 4-wire service to the nondim circuit on the switchboard in studio No. 1. There are, of course, many other circuits fed from the secondary bus transformers, for instance, a circuit to provide power to the remote

Fig. 2. View from catwalk: original lighting equipment

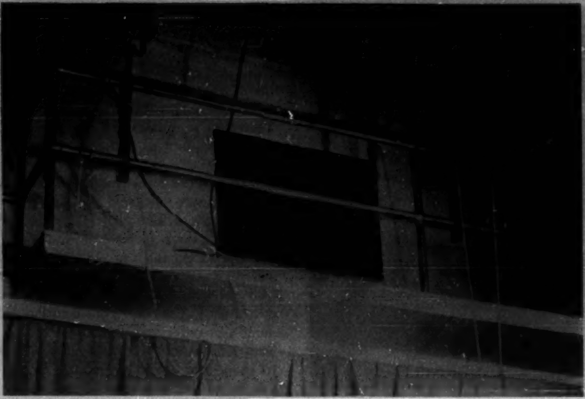


Fig. 3. Original switchboard

truck for adjusting equipment while the truck is parked in the garage.

New Lighting Switchboard

The new studio switchboard has a capacity of 1,800 amperes. The nondim circuit contains 400-ampere main fuses and a main switch. There are 12 100-ampere subbreakers. The dim circuit contains 200-ampere main fuses and a main switch. There are 12 50-ampere subbreakers each feeding a 5.5-kw Ward Leonard autrastat dimmer. Master and submaster dimmer control handles are provided.

Rotolector switches are used in the new switchboard. They are a convenient means of crossconnecting a load circuit to any supply circuit. Twenty-four-point rotolectors are used, corresponding to the 24 subcircuits, 12 dim and 12 nondim. In all, there are 140 20-ampere and 10 50-ampere units, each with a circuit breaker and pilot light. Twelve ammeters across the top meter the dimmer circuits to allow the operator to load the dimmers to their full capacity. The switchboard and rotolectors are on a platform elevated 3 feet above the studio floor. The valuable studio floor space lost is offset by the convenience of a nearby location; see Fig. 4. A tremendous saving in power has been realized from this new location. With the new lighting load it would be a very great waste to have lamps burning that were not essential. The new switchboard location makes it practical to keep the lights off unless in actual use. The turn-and-tilt controls were moved along with the switching to the platform.

Load Circuits

Under the old system the load circuits were fed from the relay contactors into the studio in 4-inch-square ducts with hinged covers. There was a total of 120 twist-lock receptacles in connecting boxes. Fortunately, there was plenty of space in these ducts to accommodate additional wiring. The load wire to each receptacle was disconnected from its relay terminal, lengthened, and connected to its proper rotolector terminal, thus utilizing most of the existing wiring and all of the existing receptacles. Under the old system a separate neutral wire was connected to three receptacles. It was necessary to add neutrals so that the neutral of each receptacle was individually fed back to the rotolector board. Twenty additional 20-ampere twist-lock and 10 50-ampere pin-type receptacles were added to these ducts above the grid.

New Fixtures

The 18 Television Associates incandescent banks in use since 1948 were retained. These continued to use both the *PAR-38* and *R-40* types of lamps on separate circuits, and retained their turn-and-tilt controls. The 15 slim-line fluorescent fixtures were taken down, not because of any adverse experience with the light quality, but because of their physical size and space limitations; see Fig. 5. Fifty 18-inch scoops were added, each using 2-kw clear *PS-52* general service lamps. Eighteen scoops were hung on individual pantagraphs. To make it possible to lower these scoops to eye level the pantagraphs were installed on a 1-inch pipe grid suspended 3 feet below the permanent grid. The remaining 32 scoops were hung on pendant clamps with a 4-foot height adjustment. These in turn were hung in pairs from turning spindles similar to the turning devices on the Television Associates telelights. These we made up from spare parts and 3/4-inch pipe. Thirty 8-inch fresnel spots were added for back and key lighting. Two-kw

Fig. 4. New switchboard; dimmers at left, rotolector switches at right

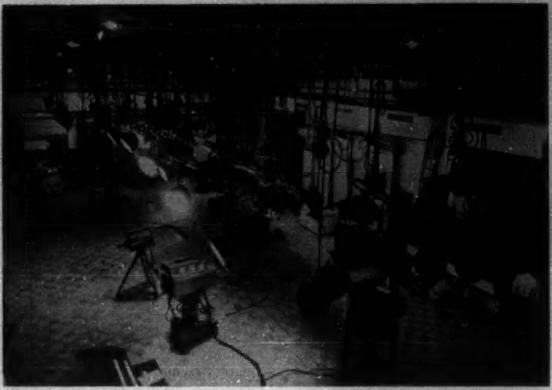


Fig. 5. New lighting layout. Note switchboard at right

G-48 mogul bipost lamps are being used in these spots. For effect use, five 16-inch fresnel spots using 5-kw mogul bipost lamps were added; for similar use, six 6-inch strips each containing 12 150-watt *R-40* spot lamps; two 750-watt boom lights; and two 8-inch 2-kw pattern projectors. There are now 60 watts per square foot as compared with 18.5 previously used; see Fig. 6.

Briefly, the considerations which prompted the types and number of fixtures, the control equipment, and the method of installation are these: First, the station was in no position to scrap its investment in the Telelight incandescent 12-lamp banks. Although these fixtures are old and the light output not as controllable as sometimes might be desired, they do provide a large quantity of light, which is one of the essentials for lighting a color set. Therefore, the first consideration was the integration of the incandescent fixtures, at least, if not the fluorescent fixtures, with the new lighting system. The second consideration was the programming. It was planned to start from the first with $5^1/_2$ hours per week of live studio programs. This made it advisable to provide one end and one side of the studio with 400 foot-candles of light at any one time. Further it was desirable to be able to swing the light quickly from one side of the studio to the other during a program. As an example of actual practice: If scene 1 is along the west wall and scene 2 is on the air from the south end of the studio, the lights can be readjusted by one man from the west side over to scene 3 along the east wall. This leads to the third consideration, that no additional personnel would be added to the staff to take care of the lighting. One man is responsible for the lighting and for lighting equipment maintenance. He has other duties, but before and during the color shows he devotes his entire time to lighting. This is the reason as many fixtures as possible have turn-and-tilt adjustments controllable from a central location. The next consideration involves accurate intensity control. It was desired to light all sets with a uniform light level of 400 foot-candles plus or minus no more than 25 foot-candles. Lighting effects, particularly those involving dimming, have always been important to the station and it was desired to continue or even expand the use of effects. Therefore, to make the lighting as uniform as possible, as well as to make possible any effects the directors might like to try, the rotolector board and the dimmers were included in the plans. In any case, a new control system was necessary, since it was physically impossible to expand the old switchboard and relay system to carry the additional load, even if the old system had been continued. The last consideration, in numerical order only, was the budget. The management was, of course, willing to spend the amount necessary to provide a good color picture. Some effects qualified, but large numbers of elaborate devices and accessories did not.

There are always new products on the market, or changes, or expansions to plan for and this station, like other TV stations, is continually ready for improvements. As soon as funds are available, plans call for conversion of another studio 30 by 40 feet, from black and white to color, additional spots for use as back light in the large studio, and possibly later conversion to pin-type cable connectors.

It is not the intention of this paper to imply that the setup described is ideal, but after nearly 2 years of color operation it has proved itself very practical for the type of operation in this studio.

Vol. 75
Pt. 1
1956

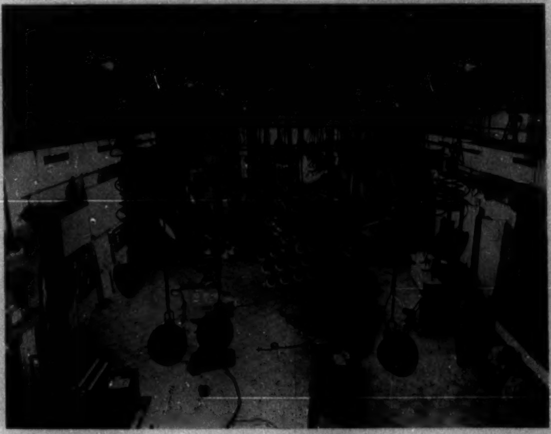


Fig. 6. New lighting layout

The Operation of the Self-Balancing Magnetic Amplifier

ALBERT D. KRALL
NONMEMBER AIEE

EDWARD T. HOOPER
ASSOCIATE MEMBER AIEE

FEEDBACK has been applied to amplifiers almost since their inception to alter the gain, bandwidth, impedances, etc. A combination of positive and negative feedbacks, decoupled from each other was applied to a vacuum-tube voltmeter by Gilbert.[1] Geyger has reported on[2] and patented[3] a magnetic circuit similar in operation to the Gilbert circuit which has been designated the self-balancing magnetic amplifier. This amplifier has appeared in many forms. Voltage and current types, both single-ended and push-pull, are possible configurations. These amplifiers have been shown, experimentally, to have large power gains, linear output-input characteristics, and a high speed of response. Scorgie[4] in comparing the self-balancing-type magnetic amplifier with the half-wave amplifier of Ramey has analyzed this circuit and reported that the speed of response can be made one half-cycle of the power supply frequency.

The purpose of this paper is two fold: first, to set forth the principle of self-balancing amplifiers; and, second, to clarify the operation of the self-balancing magnetic amplifier by means of detailed description and analysis. Since the magnetic amplifier operates on a basis of average output as a function of average input, a method of analysis will be employed that utilizes this characteristic. This particular method has been introduced to the magnetic amplifier art by Johannessen.[5]

Self-Balancing Amplifier Concept

To introduce the principle of the self-balancing amplifier, the block diagram shown in Fig. 1 will be discussed. It is not intended that the blocks represent individual parts of the self-balancing magnetic amplifier. The over-all characteristics of the block diagram are inferred to be the properties of an ideal self-balancing amplifier of any type.

A signal S is applied to the real input. It proceeds to the output after having been amplified to a value of kS. At the output there are two feedback loops. Considering first the lower or negative feedback loop, the output kS is attenuated to its original value of S and applied negatively to the real input. Thus, at the real input the originally applied signal S is completely balanced or negated. Simultaneously the upper or positive feedback loop applies a positive signal of value S to the apparent input. Isolation between the feedbacks is necessary to prevent their interaction or cancellation. The circuit is now stabilized with a finite output and zero input power. The signal source was required to supply power only during the transient. For this reason it is important to have an amplifier possessing fast response. There are four properties of this analogy which will be considered the fundamental requisites of self-balancing operation. First, the input power is zero after the initial transient; second, the output power is finite; third the signal gain is k; fourth, the response must be rapid to minimize transient power drain from the input. The first and second properties lead to a power gain approaching infinity. It must be noted that infinite power gain is a result of zero input power and not a result of high signal gain.

No restriction was made on the nature of the signal S. The self-balancing principle can be applied to balance either a current signal or a voltage signal. When a voltage input signal is balanced by feedback there results zero input current and a finite voltage gain, k. When a current input signal is balanced by feedback, there results zero input voltage drop and a finite current gain, k. It will be shown that the so-called self-balancing magnetic amplifier fulfills these requirements.

Circuit Operation

From the various types and combinations, the single-ended voltage-type self-balancing magnetic amplifier has been chosen for discussion. Assume first the familiar magnetic amplifier circuit shown in Fig. 2. This is the series-connected saturable reactor with external feedback. From this circuit the self-balancing magnetic amplifier will be evolved. As this circuit has been analyzed and discussed at length in the literature, no attempt will be made here to explain its complete operation. In general, the assumptions are made of ideal rectifiers and square-loop core material having no magnetizing current. With these assumptions Johannessen derives the following equation for the circuit of Fig. 2 where $N_F = N_L$.

$$E_0(n+1)=\frac{N_L+N_F}{N_C}E_c(n)+E_0(n)$$

The terms $E_0(n)$ and $E_0(n+1)$ are the respective average output voltages for nth and $(n+1)$th half-cycles of e_s and $E_c(n)$ is the average control voltage for the nth half-cycle. Note that the output on the $(n+1)$th half-cycle is a function of the output on the nth half-cycle. Thus a feedback is involved. The equation shows that the amplifier is unstable because the coefficient of the feedback factor $E_0(n)$ is unity. Unity feedback in this case is sometimes referred to as 100-percent postive magnetic feedback. It can also be seen from the equation that, since $E_0(n+1)$ never reaches its theoretical final value, the theoretical time constant is infinite. In a practical case the amplifier will saturate at some point and yield a finite output and time constant.

Before self-balancing can be attempted, the time constant must be reduced. This is accomplished by negating the positive feedback signal in the control circuit by the introduction of a negative electric feedback.[6] Into the circuit of Fig. 2 has been added a resistance R_K, as shown in Fig. 3. R_K is electrically common to the output and the control circuits and is the element responsible for the negative electric feedback.

The circuit of Fig. 3 will be analyzed to show the effect of R_K on the time constant and gain and to show the presence of some of the properties of an ideal self-balancing amplifier. The discussion is supplemented by the more rigorous mathematical analysis contained in Appendix I.

The basic assumptions for this analysis and discussion are as follows. The core material is considered to have a rectangular hysteresis loop, with a finite ampere-turn coercive and volt-time integration required to change flux on the loop. Ideal rectifiers have been assumed. Throughout the discussion the windings N_L are equal to the windings N_F. There are two basic modes of operation. Mode 1 is defined when both cores are unsaturated and mode 2 is defined when one core

---

Paper 56-204, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 24, 1955; made available for printing December 22, 1955.

ALBERT D. KRALL and EDWARD T. HOOPER are with the U. S. Naval Ordnance Laboratory, Silver Spring, Md.

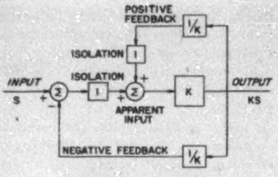


Fig. 1. Block diagram illustrating the self-balancing principle

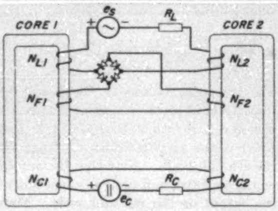


Fig. 2. Basic amplifier (series-connected saturable reactor with external feedback)

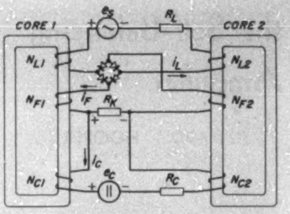


Fig. 3. Basic amplifier with negative electric feedback

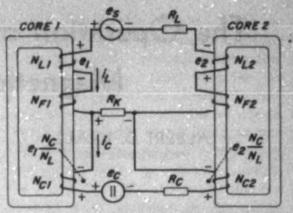


Fig. 4. Equivalent circuit of Fig. 3 for mode 1 of the nth half-cycle

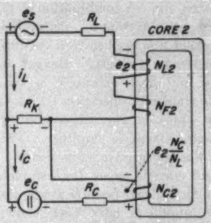


Fig. 5. Equivalent circuit of Fig. 3 for mode 2 of the nth half-cycle

is saturated and the other is unsaturated. These are the only modes permitted for proper operation. For any given half-cycle of the supply voltage e_s, one core will remain unsaturated, integrate a voltage, and flux will be set into it. The other core will have had flux set into it on the previous half-cycle; it will integrate a voltage, change the flux, and allow the core to saturate. Upon the following half-cycle the circuitry will reverse the roles previously performed by each core. For this reason only one-half cycle of operation will be analyzed; the other will be implied. Assume the source voltage e_s is of the polarity noted in Fig. 3. This particular half-cycle will be called the *n*th half-cycle. For the *n*th half-cycle the basic circuit has been redrawn, excluding the rectifier symbols, in Fig. 4. (The equivalent circuit for mode 2 is shown in Fig. 5.) Both cores start the half-cycle unsaturated. As the power supply e_s drives magnetizing current around the circuit, a voltage appears across the core windings. In core *1* the fluxes from N_F and N_L are additive; in core *2* they negate. Since N_F equals N_L, the only winding effective on core *2* is N_C. As long as core *2* remains unsaturated, the control current will be limited to the magnetizing current of N_C of core *2*. Magnetizing current will flow in the power circuit limited by core *1* and the voltage drops across the relatively low valued resistors R_L and R_K are considered negligible in comparison with other voltages in the power circuit. The voltage across N_{L1} is induced into N_{C1} of the control circuit. This induced voltage is in the order of magnitude of the supply voltage when N_L and N_C are of the same order and its polarity is shown in Fig. 4. The equation for the voltages summed around the control circuit is

$$e_1\frac{N_C}{N_L}=i_c(R_C+R_K)-i_LR_K+e_c+e_2\frac{N_C}{N_L}$$

At quiescent conditions ($e_c=0$) almost all of the induced voltage (e_1N_C/N_L) is dropped by the impedance of N_{C2}, since the resistors R_K and R_C are small in comparison to the impedance of N_{C2}. This voltage e_2N_C/N_L that is impressed on N_{C2} from N_{C1} causes a flux to be set in core *2*. Note that e_2N_C/N_L is not induced from N_{L2} because, as it was previously mentioned, the fluxes on the windings N_{L2} and N_{F2} negate each other. Depending on previous history, core *1* will or will not saturate during the *n*th half-cycle. If quiescent conditions had existed, core *1* will not saturate. Consequently, for the full *n*th half-cycle N_{C2} has a voltage impressed on it and enough flux will be set to keep core 2 from saturating on the (n+1)th half-cycle. If, however, a signal e_c exists of the polarity shown, core *1* will saturate. At the saturation of core *1* the induced voltage (e_1N_C/N_L) will drop to zero and the equation of the control circuit becomes

$$i_LR_K=i_c(R_C+R_K)+e_c+e_2\frac{N_C}{N_L}$$

Saturation of core *1* allows the load current i_L, which had previously been small, to assume a value equal to $e_s/(R_L+R_K)$. Thus in the control circuit as the induced voltage e_1N_C/N_L goes to zero, another voltage across R_K replaces it. When the value of R_K is adjusted so that the ratio $\dfrac{N_F+N_L}{N_C}$ equals the ratio $\dfrac{R_L+R_K}{R_K}$, the voltage drop across R_K exactly replaces the induced voltage e_1N_C/N_L that would have appeared across N_{C1} at quiescent conditions. The effect of this adjustment is to cause a constant voltage to appear across the combined components N_{C1} and R_K no matter what the saturation point of core *1* may be.

This constant average voltage, hereafter known as V_c, is opposed by the average control signal E_c and the net voltage V_c-E_c appears as e_2N_C/N_L across N_{C2} to set flux in core *2*. When E_c approaches V_c, a reduced amount of flux is set into core *2*. As a result of this core *2* saturates earlier in the (n+1)th half-cycle and more average load current flows. Thus, as the control signal increases from zero, the load current increased from magnetizing current. When E_c equals V_c no flux is set in core *2* and full output will occur on the (n+1)th half-cycle. Equation 19 describes the operation of the circuit.

$$E_0(n+1)=\left[\frac{N_L+N_F}{N_C}\,\frac{R_L}{R_L+R_K}\right]E_c(n)+\left[1-\frac{N_L+N_F}{N_C}\,\frac{R_K}{R_L+R_K}\right]E_0(n)$$

This equation and the equation of the series-type saturable reactor with external feedback are of the same form. The addition of R_K to the circuit has changed the coefficients of the terms. The reduction of R_K to zero in equation 19 reproduces exactly Johannessen's equation for the series-connected saturate reactor with external feedback.

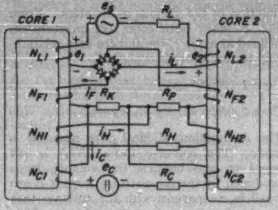


Fig. 6. Basic amplifier with negative electric feedback and "boosting" circuit

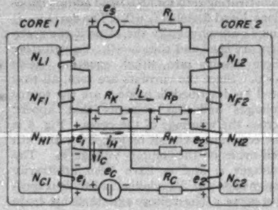


Fig. 7. Equivalent circuit of Fig. 6 for mode 1 of the nth half-cycle

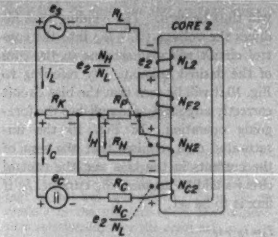


Fig. 8. Equivalent circuit of Fig. 6 for mode 2 of the nth half-cycle

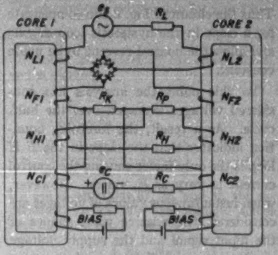


Fig. 9. Basic amplifier with negative electric feedback, "boosting" circuit, and bias

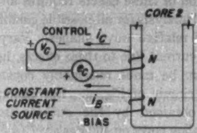


Fig. 10. Combined control and bias operation

The adjustment of the ratios $\frac{N_L+N_F}{N_C}=\frac{R_L+R_K}{R_K}$ to provide a constant V_e, as discussed previously, reduces equation 19 to

$$E_o(n+1)=\frac{R_L}{R_K}E_c(n)+[0]E_o(n)$$

This equation states that the average output voltage on the $(n+1)$th half-cycle is only a function of the average control voltage of the nth half-cycle. Thus the time constant or delay has been reduced to one half-cycle. The voltage gain has been reduced from infinity to a value equal to R_L/R_K. An increase in voltage gain greater than R_L/R_K can be obtained but only at the expense of the speed of response. To this point it has been shown that the circuit of Fig. 3 has three of the characteristics of an ideal self-balancing amplifier. It has a fast response, a finite output power, and a finite signal gain.

The fourth characteristic, that of zero input power, is so far present only if magnetizing current in the control circuit is neglected. However, for many applications the input current is required to be much lower than magnetizing current.

Before proceeding, a brief résumé of what has been done will be made. To the series-connected saturable reactor with external feedback a resistor R_K has been added to provide negative electric feedback. By making the appropriate adjustment of R_K the amplifier voltage gain is represented by a resistance ratio which in turn is equivalent to the turns ratio from the input to the output. The voltage gain here considers the output voltage as being not just composed of the voltage across R_L but of the voltage developed across the total resistance (R_L+R_K) of the output circuit. In other words, at half-cycle response the voltage amplification is the same as that of a simple transformer. As there is no feedback term present in the resulting equation, the amplifier has a gain K if the half-cycle dead time is neglected. That is, even though the "100-per-cent positive magnetic feedback" and "100-per-cent negative electric feedback" may be present, their effects cancel, resulting in no apparent feedback.

It was stated that there is always one unsaturated uncoupled winding in the control circuit. A voltage appears across this winding because of the unbalanced voltage sources in the control loop. The integral of this voltage produces a flux resetting the core. Because of the nature of an unsaturated core, only magnetizing current is allowed to flow through the control winding no matter how large the voltage across the winding is. Now it would be advantageous to reduce the magnetizing current to zero in the control circuit because that would mean zero input power.

If the control current is constant at magnetizing current or less with a changing e_c, a bias circuit, consisting of a current source on an additional winding, is sufficient to reduce the control current current to zero, since the current will then be supplied by the constant current source instead of the control circuit.

In practice when I_c is not constant with changing e_c, a circuit that resembles the conventional boosting circuit is used to produce a constant I_c. This "boosting" circuit is added to Fig. 3 to produce the circuit of Fig. 6. This circuit is analyzed in Appendix II (see Figs. 7 and 8) and the defining equation is

$$E_o(n+1)=\frac{N_L+N_F}{N_C}\frac{R_L}{R_L+R_K+R_P}E_c(n)+\left[1-\frac{N_L+N_F}{N_C}\frac{R_K}{R_L+R_K+R_P}\right]E_o(n)$$

To eliminate the feedback term in the defining equation $(N_L+N_F)/N_C$ is required to equal $(R_L+R_K+R_P)/R_K$ and whereupon the defining equation becomes

$$E_o(n+1)=\frac{R_L}{R_K}E_c(n)$$

Note that the adjustment of the ratios now accounts for the effect of the added circuit by including R_P. The circuit of Fig. 6, when the "boosting" circuit is properly adjusted, now has the same voltage gain and response time as the circuit of Fig. 3.

To make the control current constant in the circuit of Fig. 6, the derivative of the control circuit current with respect to the control voltage was equated to zero in Appendix III. The condition to be met is that

$$(R_K+R_P)\frac{di_L}{de_c}+\frac{N_L}{N_C}(R_H+R_P)\frac{dI_o}{de_c}=1$$

When this equation is satisfied, the control circuit current will be constant and can be again shifted to zero by the addition of a bias circuit. The term di_L/de_c for half-cycle response is constrained to equal $1/R_K$ since

$$R_L\frac{di_L}{de_C}=\frac{dE_o}{de_c}=\frac{R_L}{R_K}$$

Thus the circuit of Fig. 6 can compensate for a slope in the hysteresis loop dI_0/de_c and allow the control current to be made constant.

To this point the analysis has considered voltages averaged over the half-cycle. As a result, the foregoing equation indicates that the slope in the *B-H* loop is the only factor causing the control current to vary as E_c varies. However, when instantaneous values of voltages are considered, the relative wave shapes of the input signal and the supply voltage can have a great effect upon the average current flowing in the control circuit. A discussion of these effects requires an individual analysis for all possible combinations of relative wave shapes. Varying control current due to these effects, however, can be made reasonably constant by the circuit of Fig. 6. For example, values obtained by the authors using supermalloy cores on 400 cps are I_c(d-c) = ±1.5 microamperes for a control voltage range of 0.5 to 10 volts supplied by pure d-c source.

The fourth requisite for self-balancing i.e. zero input power, has now been achieved in the voltage-type amplifier shown in Fig. 9 by reducing the control current to zero for a range of control voltages. It was necessary to perform two operations simultaneously to achieve zero input current and still control the amplifier. The bias circuit performs these two operations but before showing the effects of the bias circuit in Fig. 10, the normal operation of the control circuit will be reviewed. In the control circuit there exists two voltage sources, the constant voltage V_c and the varying input signal E_c. The difference between these two voltage sources (V_c-E_c) appears across the control winding and is integrated, according to Faraday's law. The flux produced by the voltage integration varies with the input voltage and is necessary for proper operation of the amplifier. When (V_c-E_c) is applied to the control winding, the magnetic core material, assumed to have a rectangular loop, allows only magnetizing current to flow through the control winding. Magnetizing current is of constant magnitude, i.e., independent of the magnitude of (V_c-E_c). Thus as E_c varies from zero and approaches V_c, the current i_c is constant and has a value of magnetizing current I_0.

It is desirable to reduce i_c to zero but have the flux set into the core remain the same. One way to reduce i_c to zero is to induce a voltage into the control winding whose magnitude is (V_c-E_c). If this is possible, all of the voltage sources in the control circuit will sum to zero and i_c will thus be zero. The winding that induces the voltage (V_c-E_c) into the control circuit will also set flux in the core of the desired amount. By referring to Fig. 10 it will be seen that the bias circuit correctly adjusted, accomplishes the foregoing operation. In Fig. 10 the unsaturated core requires that the sum of the currents in the bias and the control circuits add to magnetizing current, I_0, if flux is to be set.

$$I_0=i_B+i_c$$

By adjusting the constant current source to produce i_B equal to I_0, the core requirements are then satisfied if i_c equals zero. Due to the nature of a constant current source adjusted to supply $i_B=I_0$, any voltage necessary to accomplish this will automatically be supplied across the bias winding. The voltage necessary to allow $i_B=I_0$ is exactly (V_c-E_c) and (V_c-E_c) will appear across the bias winding. From there it is transformed into the control winding summing all of the control voltage sources to zero and thus i_c becomes zero. The voltage (V_c-E_c) across the bias winding is also integrated, supplying the flux to the core necessary for the proper over-all circuit operation.

In practice i_c will approach zero depending on how well the "boosting" circuit can make i_c constant with E_c and on how near to ideal a constant current source is available.

With reference to the block diagram of Fig. 1, the negative feedback loop can be identified with the action of the bias circuit in reducing control current to zero. The simultaneous action of the bias circuit in setting the flux in the unsaturated core can be identified with the positive feedback into the apparent input.

Conclusion

Self-balancing amplifiers can operate from zero input power, exclusive of the transient input power, and deliver finite output power. A finite signal gain (either voltage or current) is obtained.

The self-balancing magnetic amplifier is shown to possess these properties with a speed of response that can easily be made one half-cycle of the power frequency.

The action of the "boosting" and bias circuits in relation to the self-balancing operation has been explained.

Appendix I

Fig. 3 is the diagram of the amplifier to be analyzed. The analysis will follow closely the method of Johannessen. The mathematical expressions are directed toward writing the average output voltage as a function of the various parameters. From Fig. 3, when both cores are unsaturated, defined as mode 1, the ampere-turn relationship for each core is expressed.

$$\text{Core } \mathit{1}\ (NI_0)=i_LN_L+i_FN_F-i_cN_c \tag{1}$$

$$\text{Core } \mathit{2}\ (NI_0)=i_LN_L-i_FN_F+i_cN_c \tag{2}$$

The assumption of ideal core material is made. The coercive force (NI_0) is set equal to zero because the voltage drops produced by the magnetizing current I_0 are small in comparison with the voltages that appear in the rest of the circuit. This condition implies not only that I_0 is small but that R_K, R_L, and R_C be relatively small. Substituting zero for (NI_0) and adding equations 1 and 2 yields $i_L=0$. The currents i_L and i_F are physically the same if the rectifiers are ideal and thus $i_F=0$. Substituting these results into either equation yields $i_c=0$. Since the currents are zero, all that need be described in mode 1 are the source voltages and the voltages that appear across the windings of the cores. Fig. 4 is the equivalent circuit of Fig. 3 for the *n*th half-cycle of the supply voltage e_s. The voltages e_1 and e_2 have been arbitrarily assigned to the N_L windings of their respective cores. The voltage equations for mode 1 of Fig. 4 are

$$e_s=e_1+\frac{N_F}{N_L}e_1-\frac{N_F}{N_L}e_2+e_2 \tag{3}$$

$$e_c=\frac{N_C}{N_L}e_1-\frac{N_C}{N_L}e_2 \tag{4}$$

For this analysis N_F is always equal to N_L. Substituting this condition and solving equations 3 and 4 for e_1 and e_2 yields

$$e_1=\frac{e_s}{2} \tag{5}$$

$$e_2=\frac{e_s}{2}-\frac{N_L}{N_C}e_c \tag{6}$$

Equations 5 and 6 give the voltages across the N_L windings of the reactors for mode 1 of the *n*th half-cycle. The action of the rectifier bridge will interchange the values of e_1 and e_2 during mode 1 of the $(n+1)$th half-cycle.

Mode 2 is defined when one core saturates. Core *1* can and will saturate during the *n*th half-cycle except for quiescent conditions. When core *1* saturates, the load current i_L will increase and no longer be negligible. Fig. 5 is the equivalent circuit of Fig. 4 for mode 2 of the *n*th half-cycle. The voltage equations of Fig. 5 are

$$e_s=i_L(R_K+R_L)-i_CR_K+e_2-e_2 \tag{7}$$

$$e_c=i_LR_K-i_C(R_K+R_C)-e_2\frac{N_C}{N_L} \tag{8}$$

The ampere-turn relationship for core *2* is

$$(NI_0)=i_L(N_L-N_F)+i_CN_C=i_CN_C \tag{9}$$

Again the coercive force (NI_0) is considered zero and equations 7, 8, and 9 are solved for e_2 and i_L.

$$e_2=e_s\frac{N_L}{N_C}\frac{R_K}{R_K+R_C}-e_C\frac{N_L}{N_C} \tag{10}$$

$$i_L=\frac{e_s}{R_K+R_L} \tag{11}$$

Equation 10 is the expression for the voltage across the N_L winding of core *2* during mode 2 of the *n*th half-cycle. The voltage across core *1* is zero because core *1* is saturated. Equation 11 is the expression for the load current during mode 2 and will hold true for both the *n*th and the $(n+1)$th half-cycles because of the symmetry of the circuit.

The only thing remaining before the voltage e_2 can be completely described for the *n*th and the $(n+1)$th half-cycles is a factor relating the duration of modes 1 and 2. Let the factor $a(n)$ be the ratio of the volt-time area of the supply voltage e_s during the saturation of core *1* to the total volt-time area of the supply voltage. As $a(n)$ is defined for the *n*th half-cycle, $a(n+1)$ is defined similarly for the $(n+1)$th half-cycle. From equations 6, 10, and the factor $a(n)$, the average voltage across the N_L winding of core *2* for the complete *n*th half-cycle is

$$E_2(n)=[1-a(n)]\left[\frac{E_s}{2}-\frac{N_L}{N_C}E_C(n)\right]+ a(n)\frac{N_L}{N_C}\left[E_S\frac{R_K}{R_K+R_L}-E_C(n)\right] \quad (12)$$

Similarly, using equation 5 and the factor $a(n+1)$, E_2 for the $(n+1)$th half-cycle is

$$E_2(n+1)=[1-a(n+1)]\frac{E_s}{2} \quad (13)$$

During a complete cycle of operation of any reactor that begins and ends with the same amount of flux, there is no change in the net amount of flux. If there is no change in the net amount of flux, the net average voltage across the reactor during the complete cycle is zero.

By studying the operation of core *2* for the *n*th and the $(n+1)$th half-cycles, it can be seen that the flux at the beginning of the *n*th is the same as the flux at the end of the $(n+1)$th. Thus the average voltage of core *2* over a complete cycle is zero or

$$E_2(n)=E_2(n+1) \quad (14)$$

The equations for the average output voltage across R_L for the two half-cycles are

$$E_0(n)=a(n)I_L(n)R_L=a(n)\frac{E_S}{R_L+R_K}R_L \quad (15)$$

$$E_0(n+1)=a(n+1)I_L(n+1)R_L= a(n+1)\frac{E_S}{R_L+R_K}R_L \quad (16)$$

Solving equations 15 and 16 for $a(n)$ and $a(n+1)$ yields

$$a(n)=\frac{E_0(n)}{E_S}\frac{R_K+R_L}{R_L} \quad (17)$$

$$a(n+1)=\frac{E_0(n+1)}{E_S}\frac{R_K+R_L}{R_L} \quad (18)$$

Equations 12, 13, 14, 17, and 18 are combined and solved for $E_0(n+1)$. In simplifying these equations it has been assumed that E_c is much less than E_s and that N_L is not more than an order of magnitude greater than N_C.

$$E_0(n+1)=\frac{2N_L}{N_C}\frac{R_L}{R_L+R_K}E_C(n)+ \left(1-\frac{2N_L}{N_C}\frac{R_K}{R_K+R_L}\right)E_0(n) \quad (19)$$

Equation 19 describes the average output voltage on one half-cycle as a function of the average control voltage and the average output voltage on the previous half-cycle.

Appendix II

Fig. 6 is the diagram of the amplifier to be analyzed. This analysis will differ from that of Appendix I only in that the "boosting" circuit has been added. From Fig. 6, mode 1 of the *n*th half-cycle, the ampere-turn relationships for the cores are

$$\text{Core } 1\ (NI_0)=i_LN_L+i_PN_P-i_CN_C- i_HN_H \quad (20)$$

$$\text{Core } 2\ (NI_0)=i_LN_L-i_PN_P+i_CN_C+ i_HN_H \quad (21)$$

When $N=N_L=N_P$ and $N_C=N_H$, equations 20 and 21 combine to yield

$$I_0=i_L=i_P \quad (22)$$

$$\frac{N_L}{N_C}I_0=i_C+i_H \quad (23)$$

The coercive force (NI_0) is not assumed zero until the voltage drop produced by the magnetizing current through R_H can be compared with the rest of the circuit voltages. That is, R_H is not restricted to a relatively low value as are R_K, R_L, R_C, and R_P and thus may have an appreciable voltage drop even though I_0 is small. The voltage equations for mode 1 of the *n*th half-cycle from Fig. 7 are

$$e_S=e_1+e_1+e_2-e_3+i_L(R_K+R_L+R_P)- i_CR_K-i_HR_P \quad (24)$$

$$\frac{N_C}{N_L}e_1=i_C(R_C+R_K)-i_LR_K+e_C+\frac{N_C}{N_L}e_3 \quad (25)$$

$$\frac{N_C}{N_L}e_1=i_H(R_H+R_P)+i_LR_P+\frac{N_C}{N_L}e_3 \quad (26)$$

Equations 22 through 26 are solved for e_1 and e_3 with the assumptions that e_c is at least slightly greater than zero and that R_H is much larger than $(R_P+R_K+R_C+R_L)$

$$e_1=\frac{e_S}{2} \quad (27)$$

$$e_2=\frac{e_S}{2}-\frac{N_L}{N_C}e_C \quad (28)$$

Equation 28 is the voltage across the N_L winding of core *2* during mode 1 of the *n*th half-cycle. Because of the switching action of the rectifiers, equation 27 is the voltage across the N_L winding of core *2* during mode 1 of the $(n+1)$th half-cycle.

If core *1* saturates during the *n*th half-cycle, the load current will increase. Fig. 8 is the equivalent circuit for mode 2 of the *n*th half-cycle. The voltage equations for Fig. 8 are

$$e_s=i_L(R_K+R_P+R_L)-i_CR_K-i_HR_P+ e_2-e_3 \quad (29)$$

$$i_LR_K=i_C(R_C+R_K)+e_C+\frac{N_C}{N_L}e_3 \quad (30)$$

$$i_LR_P=i_H(R_H+R_P)-\frac{N_H}{N_L}e_3 \quad (31)$$

The ampere-turn relationship for core *2* is

$$(NI_0)=i_CN_C-i_HN_H \quad (32)$$

Equations 29 through 32 are solved for i_L and e_2 yielding

$$i_L=\frac{e_S}{R_K+R_P+R_L} \quad (33)$$

$$e_2=\frac{N_S}{N_C}\left[\frac{R_K}{R_K+R_P+R_L}e_S-e_C\right] \quad (34)$$

By use of equations 27, 28, and 34, the expressions for the average voltages across core *2* for the two half-cycles are

$$E_2(n)=[1-a(n)]\left[\frac{E_S}{2}-\frac{N_L}{N_C}E_C(n)\right]+ a(n)\frac{N_L}{N_C}\left[E_S\frac{R_K}{R_K+R_L+R_P}-E_C(n)\right] \quad (35)$$

$$E_2(n+1)=[1-a(n+1)]\frac{E_S}{2} \quad (36)$$

Following the argument of Appendix I,

$$E_2(n)=E_2(n+1) \quad (37)$$

The factors $a(n)$ and $a(n+1)$ equal

$$a(n)=E_0(n)/I_L(n)R_L= \frac{E_0(n)}{E_S}\frac{R_K+R_P+R_L}{R_L} \quad (38)$$

$$a(n+1)=E_0(n+1)/I_L(n+1)R_L =\frac{E_0(n+1)}{E_S}\frac{R_K+R_P+R_L}{R_L} \quad (39)$$

The last five equations are combined in a similar manner used in Appendix I and solved for

$$E_0(n+1)=\frac{2N_L}{N_C}\frac{R_L}{R_K+R_L+R_P}E_C(n)+ \left[1-\frac{2N_L}{N_C}\frac{R_K}{R_K+R_L+R_P}\right]E_0(n) \quad (40)$$

Equation 40 describes the average output voltage on one half-cycle as a function of the average control voltage and the average output voltage on the previous half-cycle.

Appendix III

Analysis for constant current in the control circuit with changing control voltage E_C is carried out with reference to the circuit of Fig. 7. It is assumed that the windings $N_c=N_H$ in addition to the preliminary assumptions made in the other analyses. When $N_C=N_H$, the voltages across N_{C1} and N_{H1} are equal and the voltages across N_{C2} and N_{H2} are equal. Thus, in Fig. 7 the equal sets have been noted as e_1 and e_3 respectively. The voltage equations for the control and the "boosting" circuits are

$$e_1=i_C(R_C+R_K)+E_C+e_3-i_LR_K \quad (41)$$

$$e_1=i_H(R_H+R_P)+i_LR_P+e_3 \quad (42)$$

Subtracting equation 41 from equation 42 yields

$$i_H(R_H+R_P)+i_LR_P=i_C(R_C+R_K)+ E_C-i_LR_K \quad (43)$$

For the polarity of e_s, noted in Fig. 7, core *2* remains unsaturated for the entire half-cycle. The equation for the ampere-turn relationship of core *2* is

$$(NI_0)=i_L(N_L-N_F)+i_CN_C+i_HN_H \qquad (44)$$

Letting $N=N_L$ and substituting equation 44 into 43 and solving for i_C yields

$$i_C=\left[\frac{N_L}{N_C}\frac{R_H+R_P}{R_H+R_P+R_C+R_K}\right]I_0+ \left[\frac{R_P+R_K}{R_H+R_P+R_C+R_K}\right]i_L- \left[\frac{1}{R_H+R_P+R_C+R_K}\right]E_C \qquad (45)$$

It is desired that i_c remain constant with E_C. That is, the control current is to remain a constant value of direct current for a range of control voltage. When and if this condition can be accomplished, it would be relatively easy to add another winding supplied by a constant current source to each core and reduce the control current to zero.

To find the conditions necessary for a constant i_c with respect to E_c, differentiate i_c in equation 45 with respect to E_C and set the differential equal to zero. This yields

$$1=\frac{N_L}{N_C}(R_H+R_P)\frac{dI_0}{de_c}+(R_P+R_K)\frac{di_L}{de_c} \qquad (46)$$

where $\frac{dI_0}{de_c}$ is the change in magnetizing current with respect to the control voltage caused by the sloping of the hysteresis loop and $\frac{di_L}{de_c}$ is the change in load current for a change in control voltage. For one-half-cycle response $\frac{di_L}{de_c}=\frac{dE_0/de_c}{R_L}=\frac{1}{R_K}$ from equation 40 in Appendix II. Under these conditions equation 46 becomes

$$-\frac{dI_0}{de_c}=\frac{N_C}{N_L}\frac{R_P}{R_H+R_P}\frac{1}{R_K} \qquad (47)$$

Equation 47 shows the adjustment of R_P and R_H necessary to compensate for a negative slope of the hysteresis loop. When equation 47 is satisfied, the amplifier of Fig. 7 will have a constant control current for changing control voltages and a response time of one half-cycle of the power supply frequency.

References

1. Stable Voltmeter. R. W. Gilbert. *Electronics*, New York, N. Y., Mar. 1947, pp. 130–33.

2. Magnetic Amplifiers of the Self-Balancing Potentiometer Type, W. A. Geyger. *AIEE Transactions*, vol. 71, pt. I, Jan. 1952, pp. 383–95.

3. Self-Balancing Magnetic Amplifier, W. A. Geyger. *United States Patent No. 2,700,130*, Jan. 18, 1955.

4. Fast Response with Magnetic Amplifiers, D. G. Scorgie. *AIEE Transactions*, vol. 72, pt. I, Jan. 1954, pp. 741–49.

5. Analysis of Magnetic Amplifiers by the Use of Difference Equations, P. R. Johannessen. *Ibid.*, vol. 74, pt. I, Jan. 1955, pp. 700–11.

6. Magnetic Amplifiers (book). H. F. Storm. John Wiley and Sons, Inc., New York, N. Y., 1955.

Representation of Nonlinear Characteristics by Linear Elements

M. S. M. ABOU-HUSSEIN
ASSOCIATE MEMBER AIEE

THE solution of the modern complicated electrical engineering problems needs simpler methods than dealing with a large number of equations in a large number of unknowns; simpler experimental solutions are introduced in the form of simple equivalent circuits or a collection of such circuits with variable constants (to represent more than one case) and measuring instruments in a suitably designed machine called analogue computer. Hence the importance of facilitating the solution of nonlinear characteristics of magnetic circuits, neglecting hysteresis, by the steady-state response of linear elements and using the ordinary measuring instruments (ammeter and voltmeter) is clear.

In this paper it is explained how the magnetization curve of a d-c loaded magnetic circuit may be represented by the steady-state response of linear elements. The operation of the circuit is simple; one input and one output voltage are related together with the nonlinear equation. Only the output voltage is measured, for a known input voltage after a current meter reading is adjusted, with the use of two identical mechanically coupled variable resistances, to a fixed constant value.

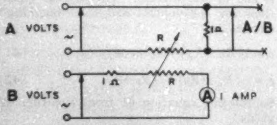


Fig. 1. Obtainment of A/B without measuring A or B

Theory

Before dealing directly with the problem of representation, the following method is given explaining how linear elements may be used to get the value A/B if both A and B are given as isolated voltages though not of known magnitudes and are obtained from the same source (having the same frequency) but through isolating voltage transformers giving output in-phase voltages of magnitudes A and B. Considering Fig. 1 and adjusting the two identical mechanically coupled resistances to the value R corresponding to the flow of 1 ampere in the ammeter

$$B=R+1$$

or

$$R=B-1 \qquad (1)$$

In the form of current, the current in the upper circuit under these conditions will be

$$A/(R+1)=A/B \qquad (2)$$

In the form of voltage, the voltage drop across the 1-ohm resistance of the upper circuit will also be A/B.

In the following representations, using some typical equations of magnetization, transformers are used as voltage multipliers; for a high degree of accuracy of operation of the equation analogue these transformers are to have negligible exciting current and regulation.

FRÖHLICH'S EQUATION ANALOGUE CIRCUITS

Case 1

$y=ax/(b+x)$; x is known, find y (3)

The analogue circuit is shown in Fig. 2 There are two isolated voltage sources of magnitudes x and b and a transformer of ratio a, so that voltages of magnitudes ax and $(b+x)$ are obtainable and divided as indicated in Fig. 1.; this division is equal to y.

Case 2

Fröhlich's equation may be reduced to the form

Paper 56-121, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted March 22, 1954; made available for printing November 21, 1955.

M. S. M. Abou-Hussein is with the University of Cairo, Giza, Egypt.

Vol.
Pt. 1
1956

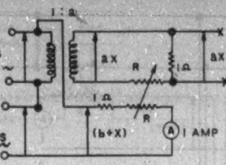
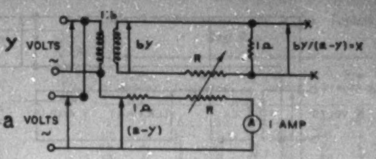


Fig. 2 (left). Solution of Fröhlich's equation: x is known, find y; $y=ax/(b+x)$

Fig. 3 (right). Solution of Fröhlich's equation: y is known, find x; $x=by/(a-y)$



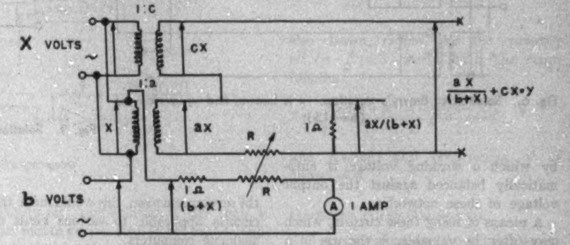


Fig. 4. Solution of modified Fröhlich's equation: x is known, find y; $y=[ax/(b+x)]+cx$

$$x=yb/(a-y);\ y \text{ is known, find } x \qquad (4)$$

which is the same equation as for case 1 but with x, y, a, and b replacing y, x, b, and a respectively and with a subtraction sign instead of an addition sign for case 1; Fig. 3 will be the required analogue circuit for this case.

Modified Fröhlich's Equation Analogue Circuits

Case 1

$$y=[ax/(b+x)]+cx;\ x \text{ is known, find } y \qquad (5)$$

Fig. 4 is the analogue circuit for this case; this figure is the same as Fig. 2 but with a second transformer circuit, so that a voltage of magnitude of cx is obtained and added to the voltage $ax/(b+x)$ across the 1-ohm resistance of the medium circuit; the sum being y.

Case 2

The modified Fröhlich's equation may be transferred to the form

$$y'=y-cx=ax/(b+x)$$

or

$$x=y'b/(a-y');\ y \text{ is known, find } x \qquad (6)$$

which is the same equation as case 2 of Fröhlich's equation so that Fig. 3 may be used but with y' replacing y and with an additional circuit giving the value of cx which is added to y'; the sum being y which is parallelled to the final input voltage, as shown in Fig. 5.

Emery's Equation Analogue Circuits

Using the same principles of voltage addition, subtraction, and division, Emery's equation of saturation of magnetic circuits could be represented in a similar way, as shown in Fig. 6, for the equation

$$y=k-[k/(mx+Lk)],\ x \text{ is known, find } y \qquad (7)$$

In Fig. 7 Emery's equation is shown for case 2 where the equation is transferred to the form

$$x=(k/m)[(k-y)^{-1}-L] \qquad (8)$$

Notes:

1. For cases 1, if x is known and changed to another value, the effect of saturation according to the set-up equation will be considered by varying the resistance R until the ammeter reads 1 ampere; then the indicated value of y will be the new value.

2. For cases 2, note 1 is repeated, replacing x by y and y by x.

3. The directions of x, y, and the applied voltages could be fixed by assuming one constant direction for positive values and the reverse direction for negative ones.

4. Direct current may be used in the analogue circuits instead of the alternating current by choosing appropriate values of potential dividers replacing transformers, the ratios of which are rendered to unity or less by representing the equation constants and variable quantities to scale in the analogue circuits.

5. It is important to notice that only d-c loaded magnetic circuits are considered in the previous representations.

6. The frequency value that may limit the operation of the circuits is either the value which makes the transformer series impedance comparable to $(R+1)$ or that value which causes resonance between the circuit inductances and stray capacitances at high frequencies.

Conclusions

It is thus possible to represent nonlinear equations of the form given by simple linear electric analogues. The accuracy of operation of the circuits is dependent on the degree of perfectness of the elements used, the method of their loading, the degree of supression of the effects of the transformer magnetizing circuits, and the supply harmonics and nonconstancy. Test results might have been given but the accuracy will vary from one type of equipment to another, and results given for one type of equipment will not apply to another. Furthermore, this type of inaccuracy is well known.

Using compensated transformers throughout the analogues, and with the existing tendency toward the use of automatic devices, it would be possible to arrange accurately automatic means to maintain the adjustments required in the circuits and to provide a balancing means

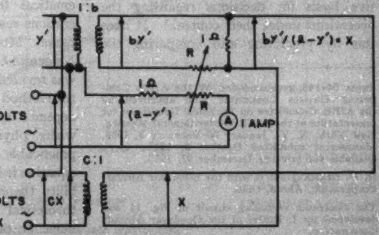


Fig. 5. Solution of modified Fröhlich's equation: y is known, find x; $y'=y-cx=ax/(b+x)$.

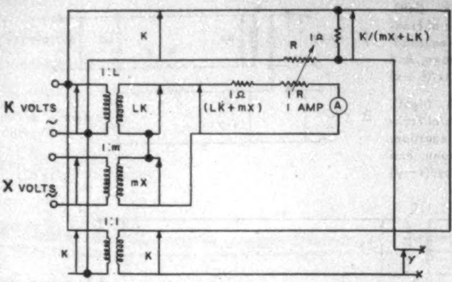


Fig. 6. Solution of Emery's equation: x is known, find y, $y = k - [k/(mx+Lk)]$

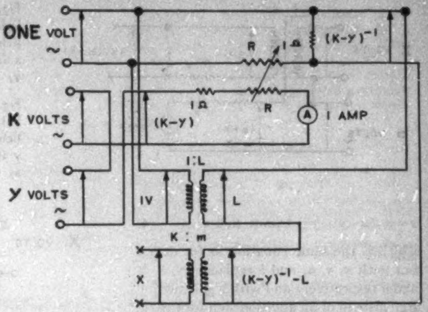


Fig. 7. Solution of Emery's equation: y is known, find x, $x = (k/m)[1/(k-y)-L]$

by which a working voltage is automatically balanced against the output voltage of these networks.

A means of using these circuits, which may often be valuable, is the use of a servomechanism system to position the potentiometer automatically. Within the limits of the speed of response on the servomechanism, this would make the circuits applicable to various kinds of analogue computers.

Hence the d-c loaded magnetic cores, which are often included in the many machines, instruments, apparatus, etc., would be thus accurately represented in a simple way with the use of linear elements, taking saturation into account, by feeding the given circuits into different analogue computers which include such electromagnetic equipment.

Analogue Computer Generation of Probability Distributions for Operations Research

N. D. DIAMANTIDES
ASSOCIATE MEMBER AIEE

OPERATIONS research is being applied in science, commerce, and industry. Developed in response to military needs during World War II, it was then known as operations analysis and later as weapons systems evaluation. In its most general aspect operations research is a scientific method of providing executive departments with a quantitative basis for decisions regarding the operations under their control.[1] It does this by predicting and comparing the values, effectiveness, and costs of alternative courses of action involving man-machine systems.

Monte Carlo Method

The most important single mathematical tool of operations research is probability and statistical theory.[1] One mathematical technique used in many operations research problems was given the name "Monte Carlo Method" by von Neumann,[2] during World War II when it was revived at Los Alamos. In general, the method is used to solve problems that depend upon probability interference where physical experimentation is impracticable and the creation of an exact equation is impossible. While the probability theory derives its equations from theoretical factors based on the phenomenon of chance, in the Monte Carlo method an answer is sought with the use of probability to a physical question often having no relation to probability except as a tool. The technique is based on the construction of a mathematical model. The model describes the behavior of related numbers assumed to bear a correspondence to events associated with physical reality. The numbers are selected at random in accordance with the statistical character of the input quantities to the model, but the operations imposed on them follow the rules of the model. Thus, the behavior of the output variable of the particular problem can be obtained by means of calculations prescribed by the model.

Probability Distribution Generation

RANDOM-NOISE GENERATOR

Application of analogue computers to operations research problems by use of the Monte Carlo method requires a source of the probability distributions most frequently encountered in applied science. One source is an electronic random-noise generator utilizing the noise component of a thyratron plate voltage kept in a condition of constant discharge. The output component is subjected to a 2-stage amplification, Fig. 1, after which the pulse peaks trigger a blocking oscillator. The output of the oscillator is a

---

Paper 56-148, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 3, 1955; made available for printing December 27, 1955.

N. D. DIAMANTIDES is with the Goodyear Aircraft Corporation, Akron, Ohio.

The electronic switching circuit of Fig. 11 was developed by J. Butler of the Goodyear Aircraft Corporation.

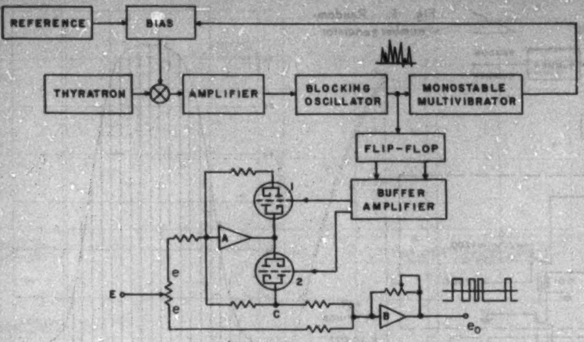


Fig. 1. Random-pulse generator

Table I. Measurement of n for Typical Test Run

n	Frequency	n	Frequency	n	Frequency	n	Frequency
9	1	15	18	21	140	27	18
10	2	16	56	22	114	28	11
11	2	17	80	23	91	29	4
12	4	18	109	24	63	30	2
13	9	19	133	25	60	31	1
14	15	20	138	26	33		

series of identical pulses with the pulse rate obeying a Gaussian distribution. By means of a monostable multivibrator following the oscillator stage, the grid bias of the first amplification stage is adjusted automatically to ensure a constant mean counting rate. The output of the oscillator, acting through a flip-flop and a buffer amplifier, switches in opposite directions the two parallel feedback impedances of another d-c amplifier *A*. When the electronic switch 1 is "on," switch 2 is "off," point *C* is at zero potential, and the only input to output amplifier *B* is the reference voltage $+e$. When switch 1 is off and switch 2 is on, point *C* is at $-2e$ potential and the effective input to amplifier *B* is $-e$. A symmetric square-wave output, e_0, results with the duration of the alternate pulses being governed by the Gaussian distribution of the oscillator output. The noise generator can be combined with analogue computer operational units to generate not only Gaussian but other random distributions.

Gaussian Amplitude Distribution

The number n of the blocking oscillator output pulses per 0.33 second was measured for a typical test run. The results, as shown in Table I, were used to check the adequacy of fit of the pulse-rate frequency. An indication of whether the Gaussian distribution applies can be obtained by making use of the so-called probability chart. The abscissa scale of this chart is nonuniform, following the law of normal probability. If the plot of n versus percentage of cases is a straight line, it can be assumed that the phenomenon does obey the normal law. Engineering phenomena cannot be counted on to follow the normal law closely into probabilities of less than 1 per cent or greater than 99 per cent.[3] The pulse rate in this case shows a satisfactory degree of Gaussian distribution; see Fig. 2. The mean counting rate is $N=20$ pulses per 0.33 second and the standard deviation is $\sigma=3.2$ pulses per 0.33 second. Thus, the probability distribution will be

$$P(n)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(n-N)^2}{2\sigma^2}} \tag{1}$$

$$P=0.124e^{-\frac{(n-20)^2}{20.48}} \tag{2}$$

N is determined by the ordinate corresponding to the value 50 of the percentage of cases; σ is given by the difference of N and the ordinate at 84.13 per cent. In the same plot, the standard-error band is also drawn, defined by the quantity erf $(n-N)\pm S$, where erf is the error-function

$$S=\sqrt{\frac{[\operatorname{erf}(n-N)][1-\operatorname{erf}(n-N)]}{Z}}$$

and Z = the total occurrences of all values of n.

Gaussian Pulse Generator

One convenient way for obtaining a sequence of random numbers of Gaussian magnitude distribution for use in place of the standard tables required by digital computers is by generating a train of identical pulses of rate n and summing them over a selected constant interval. As shown in Fig. 3, a properly limited differentiator and a rectifier, both in series with the noise generator, will provide a pulse at each zero-crossing of the square-wave output. Summing these pulses over the unit time interval will result in a total numerically equal to the pulse rate. At the end of such intervals the output of the summer will attain values which, like the pulse rate itself, will be Gaussian.

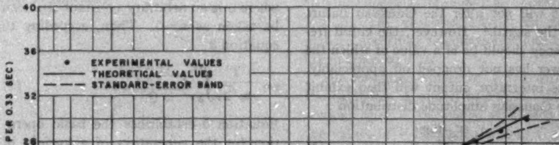


Fig. 2. Pulse-rate distribution

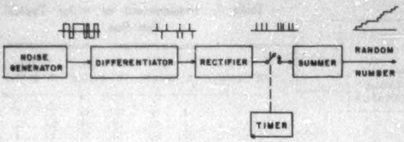

NOISE GENERATOR
DIFFERENTIATOR
RECTIFIER
S
SUMMER
RANDOM NUMBER
TIMER

Fig. 3. Random-number generator

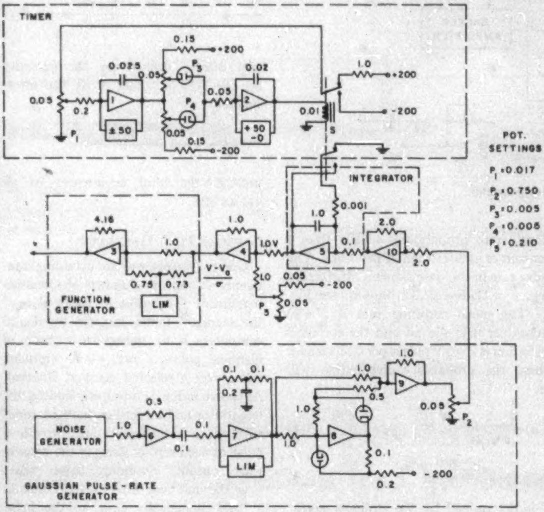

TIMER
0.15
+200
0.025
0.05
P3
0.02
1.0
+200
0.05
0.2
1
±50
P1
P4
2
+50 -0
0.01
-200
S
0.05
0.15
-200
INTEGRATOR
0.001
1.0
POT. SETTINGS
P1 = 0.017
P2 = 0.750
P3 = 0.005
P4 = 0.005
P5 = 0.210
v
4.16
3
1.0
0.75
0.73
LIM
FUNCTION GENERATOR
(V-V0)/σ
4
1.0V
5
0.1
10
2.0
2.0
1.0
0.05
-200
P5
0.05
0.1 0.1
0.2
1.0
9
0.5
0.05
P2
0.3
NOISE GENERATOR
1.0
6
0.1
0.1
7
1.0
8
LIM
0.1
0.2
-200
GAUSSIAN PULSE-RATE GENERATOR

Fig. 4. Random-distribution generator

A continuous output can be obtained by filtering the noise-generator output through a sharp cutoff low-pass filter; this will not alter the Gaussian nature of the signal.[4] However, the cutoff frequency should be of a value at which the power has not dropped off appreciably. The integrator output will then exhibit the Gaussian amplitude distribution

$$P(V)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(V-V_0)^2}{2\sigma^2}} \tag{3}$$

Other Amplitude Distributions

The most important feature of the Gaussian pulse generator is its ability to create other random distributions. Assume that a second randomly varying voltage $v=f(V)$, is sought, the instantaneous values of which will show a one-to-one correspondence to the random values of V. The probability $Q(v)$ that the voltage v lies in the interval $(v, v+dv)$ must equal the probability $P(V)$ that voltage V lies in the interval $(V, V+dV)$. Consequently,[5] $Q(v)dv=P(V)dV$

$$\text{or } v=\int_a^V \frac{P}{Q}dV \tag{4}$$

where a is an arbitrary constant that can be taken as $a=0$. Then, according to equation 3

$$v=\frac{1}{\sigma\sqrt{2\pi}}\int_a^V \frac{e^{-\frac{(V-V_0)^2}{2\sigma^2}}}{Q}dV \tag{5}$$

Equation 5 establishes the basic operation for the generation of a variety of random distributions as shown by the examples in the Appendix. Besides the ones listed there, the linear, hyperbolic, inverse quadratic, and inverse Gaussian distributions were also evaluated.

Circuitry for Probability Transformation

To obtain any of the distributions described in the Appendix, it will be necessary to implement the pertinent equation on the computer. This will require one function generator following the Gaussian source of V. For every value V, then,

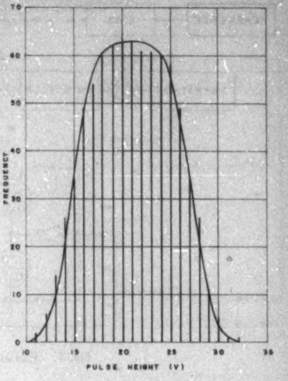


Fig. 5. Gaussian number distribution, measured

the corresponding voltage v will be generated, with its amplitude obeying the selected distribution $Q(v)$. As an example, the generation of a "flat" or uniform distribution was tried as required by equation 12. Fig. 4 shows the details of the computer circuit. The upper part of the diagram is a timer operating a sampling switch S. With the setting shown, the switch stays open for 1 second and closes for about 0.05 second to discharge the random-number integrator. At the lower part of the diagram the generator square wave is differentiated, limited to ensure constant pulse height, and rectified. The resulting pulse trains are summed by the random-number integrator. Fig. 5 shows both the experimental and theoretical data for a mean counting rate of nine counts per clock pulse. The number magnitude V is given in volts and the mean V_0 and the standard deviation σ were calculated as $V_0=21$ volts, and $\sigma=5.2$ volts, by means of a probability plot similar to that of Fig. 2. Consequently

$$P(V)=0.077e^{-\frac{(V-21)^2}{54.08}} \tag{6}$$

Generally, the distribution will be of the form given by equation 1. Since the uniform distribution is sought, application of equation 13 gives

$$v=\frac{1}{k\sigma\sqrt{2\pi}}\int_0^V e^{-\frac{(V-V_0)^2}{2\sigma^2}}dV \tag{7}$$

With

$$\frac{V-V_0}{\sigma}=x \tag{8}$$

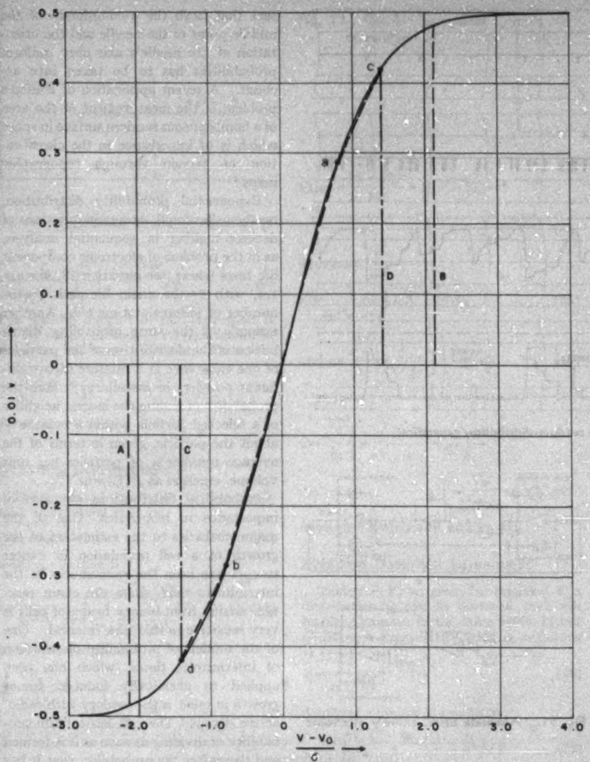


Fig. 6. Function generator characteristics

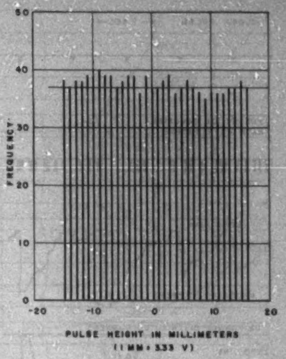


Fig. 8. Uniform number distribution, measured

equation 7 becomes

$$v=\frac{1}{k\sqrt{2\pi}}\int_{-\frac{V_0}{\sigma}}^{\frac{V-V_0}{\sigma}}e^{-\frac{x^2}{2}}dx \qquad (9)$$

$$\text{or } v=\frac{1}{k\sqrt{2\pi}}\int_{-\infty}^{\frac{V-V_0}{\sigma}}e^{-\frac{x^2}{2}}dx-$$

$$\frac{1}{k\sqrt{2\pi}}\int_{-\infty}^{-\frac{V_0}{\sigma}}e^{-\frac{x^2}{2}}dx \qquad (10)$$

This yields[1]

$$v=\frac{1}{k}\left[\operatorname{erf}\left(\frac{V-V_0}{\sigma}\right)-\operatorname{erf}\left(-\frac{V_0}{\sigma}\right)\right] \qquad (11)$$

On the basis of a table for erf(x) and of the values of V_0 and σ already derived, the function v is obtained as shown in Fig. 6. The abscissas have been normalized according to equation 8, and k is taken equal to 0.01.

The measured values of V fall within the ordinates A and B of Fig. 6. It is obvious that if the range is limited to the values of V between the lines C and D, i.e., $28.8>V>13.2$, then the v function can be approximated by the three straight-line segments *ca-ab-bd* without appreciable loss of accuracy.

The simulation circuit for the generation of the function *c-a-b-d* is shown in Fig. 7, along with the required settings. The computations were based on the assumption that one σ is represented by 5.2 volts in the function generator input. Fig. 8 gives the results of one run (1,126 numbers); Fig. 9 is a small portion of the run. A second application of the method was tried, the aim being the generation of the inverse Gaussian distribution. The results are shown in Fig. 10.

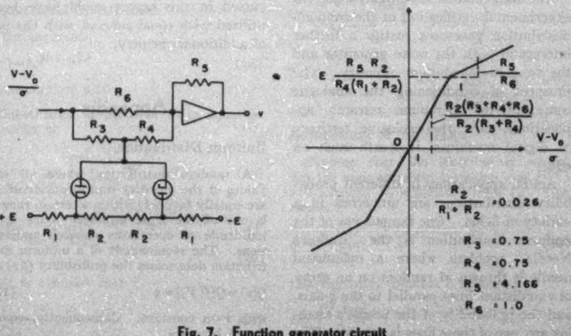


Fig. 7. Function generator circuit

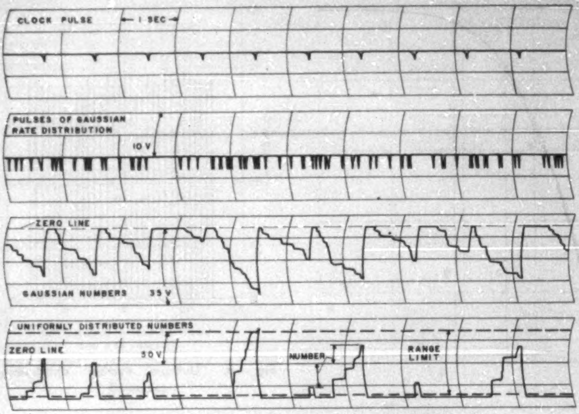


Fig. 9. Portion of run for uniform random-distribution generation

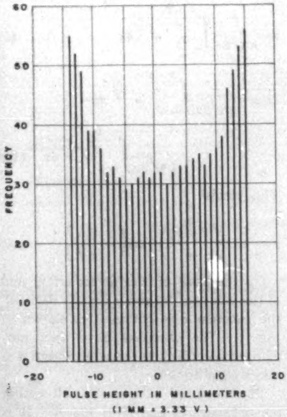


Fig. 10. Inverse Gaussian number distribution, measured

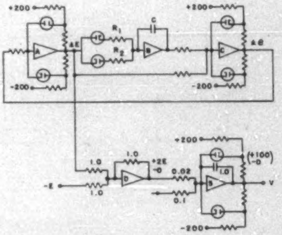


Fig. 11. Alternate timer, entirely electronic

Note on Relay in Sampling Circuit

A relay in the sampling circuit could be objectionable when higher speeds are required. In this case the purely electronic switching circuit shown in Fig. 11 can be utilized. The output of amplifier A is a square wave $\pm E$, with the duration of the halves adjusted by the time constants R_1C and R_2C (amplifier B). The addition of a $-E$ voltage in amplifier D results in equally spaced pulses $+2E$. When these pulses are fed (through an impedance assuring high gain) into the integrator that produces the Gaussian numbers V, they cancel the integrator output and prepare the integrator for the next number.

Discussion

The mathematical investigation and the experimental verification of the random-distribution generator justify a further interest in both the noise generator and the random-distribution circuitry. The prospect of conditioning the analogue computer to operations research applications, now the exclusive territory of digital machines, is worth such an effort.

Actual applications of different probability distributions are numerous in a variety of fields. One famous use of the uniform distribution is the "Buffon's Needle" problem where a cylindrical needle is thrown at random on an array of equidistant lines parallel to the x-axis, and the probability of the needle's crossing any one of these lines is sought. The fact that both the y-co-ordinate of the middle point of the needle and the orientation of the needle's axis obey uniform probabilities has to be taken into account. A recent application of Buffon's problem is the measurement of the area of a homogeneous random surface in space which is of importance in the classification of terrain through topographic maps.[6]

Exponential probability distribution, on the other hand, occurs in questions of decision-making in sequential analysis, as in the problem of electronic component life tests where, see equation 18, $k_6=a=1/\theta$, with θ=the mean life and V=the number of failures at time t.[7,8] Another example of the same probability distribution is the distribution of the particles of one substance in a mixture of two different powders in metallurgy;[9] here the probability of finding the nearest neighbor of a selected particle within a volume V about the particle, given in terms of the average number n of particles per unit volume, emerges as $P(V)=n\epsilon^{-Vn}$.

Probability distributions are also of importance in biophysics. One of the major obstacles to the estimation of the growth of a cell population in cancer tumors has been the randomness of the intermitotic time, since the chain reaction arising from binary fission of cells is very sensitive to this time interval. One of the models of probability distribution of intermitotic times, which has been applied to chemically induced tumor growth in mice, is the analogy with radioactive decay. The cell has a finite probability of dividing as soon as it is formed and thereafter the probability that it has not divided decreases exponentially. At the Los Alamos Scientific Laboratories, studies were conducted lately on this problem using Monte Carlo methods implemented by the MANIAC digital computer.[10] Obviously the technique discussed in this report could have been utilized with equal success, with the use of additional circuitry.

Appendix

Uniform Distribution

A random distribution where all the values of the quantity under consideration are equally favored (within a certain range) is one of the most important ones in a multitude of operations research applications. The requirement of a uniform distribution determines the probability $Q(v)$ as

$$Q(v)=Q[f(V)]=k \tag{12}$$

with k=a constant. Consequently, equation 5 becomes

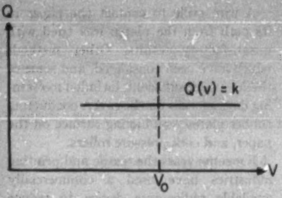


Fig. 12. Erf, definite integral of equation 13

$$v=\frac{1}{k\sigma\sqrt{2\pi}}\int_0^V e^{-\frac{(V-V_0)^2}{2\sigma^2}}\,dV \tag{13}$$

The definite integral of this expression is the erf which, because of its significance in statistics, has been tabulated. Thus

$$v=\frac{1}{k}\left[\operatorname{erf}\left(\frac{V-V_0}{\sigma}\right)-\operatorname{erf}\left(\frac{a-V_0}{\sigma}\right)\right] \tag{14}$$

shown graphically in Fig. 12.

Quadratic Distribution

The analytical expression for the quadratic distribution will be

$$Q(v)=k_2(V-V_0)^2 \tag{15}$$

making equation 5

$$v=\frac{1}{k_2\sigma\sqrt{2\pi}}\int_{V_a}^{V}\frac{e^{-\frac{(V-V_0)^2}{2\sigma^2}}}{(V-V_0)^2}\,dV \tag{16}$$

under the condition

$$\frac{V-V_0}{V_a-V_0}>1$$

$$\text{or } v=\frac{1}{k_2\sigma}\left[\operatorname{erf}\left(\frac{V-V_0}{\sigma}\right)-\operatorname{erf}\left(\frac{V_a-V_0}{\sigma}\right)\right]+$$

$$\frac{1}{k_2\sigma\sqrt{2\pi}}\left[\frac{e^{-\frac{V_a^2}{2\sigma^2}}}{V_a-V_0}-\frac{e^{-\frac{V^2}{2\sigma^2}}}{V-V_0}\right] \tag{17}$$

shown graphically in Fig. 13, which is a type-II Pearson distribution.[11]

Exponential Distribution

The expression

$$Q(v)=k_6e^{a(V-V_0)} \tag{18}$$

where $a>0$ or $a<0$.

Substitution of equation 16 in equation 5 results in

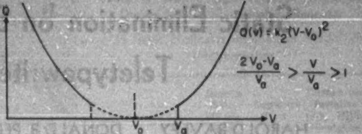
$$v=\frac{1}{k_6\sqrt{2\pi\sigma}}\int_0^V\frac{e^{-\frac{(V-V_0)^2}{2\sigma^2}}}{e^{a(V-V_0)}}\,dV \tag{19}$$

$$\text{or } v=\frac{1}{k_6\sigma\sqrt{2\pi}}\int_0^V e^{-\frac{(V-V_0)^2}{2\sigma^2}-a(V-V_0)}\,dV \tag{20}$$

If it is assumed that

$$\frac{y}{\sqrt{2}\sigma}=\frac{V-V_0}{\sqrt{2}\sigma}+\frac{a\sigma}{\sqrt{2}} \tag{21}$$

Fig. 13. Graphic expression of equation 17



$$\text{then } -\frac{(V-V_0)^2}{2\sigma^2}-a(V-V_0)=$$

$$-\frac{y^2}{2\sigma^2}+\frac{a^2\sigma^2}{2} \tag{22}$$

Therefore

$$v=\frac{1}{k_6\sigma\sqrt{2\pi}}\int e^{\frac{a^2\sigma^2}{2}-\frac{y^2}{2\sigma^2}}\,dy \tag{23}$$

$$\text{and } v=\frac{e^{\frac{a^2\sigma^2}{2}}}{k_6}\,\frac{1}{\sigma\sqrt{2\pi}}\int e^{-\frac{y^2}{2\sigma^2}}\,dy \tag{24}$$

This results in

$$v=\frac{e^{\frac{a^2\sigma^2}{2}}}{k_6}\left[\operatorname{erf}\left(\frac{V-V_0}{\sigma}+a\sigma\right)-\right.$$

$$\left.\operatorname{erf}\left(-\frac{V_0}{\sigma}+a\sigma\right)\right] \tag{25}$$

expressed graphically in Fig. 14.

Standard Deviation Adjustment

Taking $P(V)$ as given by equation 3, a distribution Q can be obtained that will be also Gaussian of the same mean V_0 but of different standard deviation σ_0, as follows

$$Q=\frac{1}{\sigma_0\sqrt{2\pi}}e^{-\frac{(V-V_0)^2}{2\sigma_0^2}} \tag{26}$$

Applying equation 5 results in

$$v=\frac{\sigma_0}{\sigma}\int_0^V e^{\frac{(V-V_0)^2}{2}\left(\frac{1}{\sigma^2}-\frac{1}{\sigma_0^2}\right)}\,dV \tag{27}$$

$$v=\frac{\sigma_0}{\sigma}\int_0^V e^{\frac{-\frac{(V-V_0)^2}{2\sigma^2\sigma_0^2}}{\sigma_0^2-\sigma^2}}\,dV \tag{28}$$

Consequently

$$v=\frac{\sigma_0^2\sqrt{2\pi}}{\sqrt{\sigma_0^2-\sigma^2}}\left[\operatorname{erf}\left(\frac{V-V_0}{\frac{\sigma_0\sigma}{\sqrt{\sigma_0^2-\sigma^2}}}\right)-\right.$$

$$\left.\operatorname{erf}\left(\frac{-V_0}{\frac{\sigma_0\sigma}{\sqrt{\sigma_0^2-\sigma^2}}}\right)\right] \tag{29}$$

Adjustment of the Mean

Suppose that the distribution sought has the same standard deviation σ as $P(V)$ but its mean V_n is different from V_0. This means

$$Q=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(V-V_n)^2}{2\sigma^2}} \tag{30}$$

Consequently, equation 5 becomes

$$v=\int_0^V e^{-\frac{1}{2\sigma^2}[(V-V_0)^2-(V-V_n)^2]}\,dV \tag{31}$$

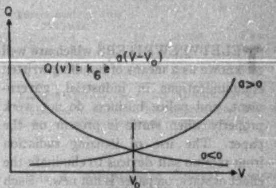


Fig. 14. Graphic expression of equation 25

$$\text{or } v=\int_0^V e^{\frac{V_n^2-V_0^2}{2\sigma^2}-\frac{V_n-V_0}{\sigma^2}V}\,dV \tag{32}$$

$$\text{and } v=e^{\frac{V_n^2-V_0^2}{2\sigma^2}}\int_0^V e^{\left(\frac{V_n-V_0}{\sigma^2}\right)V}\,dV \tag{33}$$

which after the integration becomes

$$v=\frac{\sigma^2}{V_n-V_0}e^{\frac{V_n^2-V_0^2}{2\sigma^2}}\left[1-e^{\left(\frac{V_n-V_0}{\sigma^2}\right)V}\right] \tag{34}$$

References

1. METHODS OF OPERATIONS RESEARCH (book), G. E. Kimball, P. M. Morse. John Wiley & Sons, Inc., New York, N. Y., 1951.

2. NUMERICAL SOLUTION OF DIFFERENTIAL EQUATIONS (book), W. E. Milne. John Wiley & Sons, Inc., New York, N. Y., 1953, p. 254.

3. PRACTICAL ENGINEERING USES OF PROBABILITY PAPER, A. Boyajian. *General Electric Review*, Schenectady, N. Y., June 1950, p. 11.

4. THE DISTRIBUTION OF AMPLITUDE WITH TIME IN FLUCTUATION NOISE, V. D. London. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Feb. 1941, p. 50.

5. PROBABILITY AND INFORMATION THEORY, WITH APPLICATIONS TO RADAR, P. M. Woodward. McGraw-Hill Book Company, Inc., New York, N. Y., 1953, p. 13.

6. A MEASURE OF THE AREA OF A HOMOGENEOUS RANDOM SURFACE IN SPACE, S. Corrsin. *Quarterly of Applied Mathematics*, Brown University, Providence, R. I., Jan. 1955, p. 404.

7. A SEQUENTIAL TWO SAMPLE LIFE TEST, B. Epstein. *Journal*, Franklin Institute, Philadelphia, Pa., July 1955, pp. 25–29.

8. CONTRIBUTION TO SEQUENTIAL ANALYSIS, M. A. Girslick. *Annals of Mathematical Statistics*, Michigan State College, E. Lansing, Mich., vol. 17, 1946, p. 123.

9. ANALYSIS OF CONTINUITY OF ONE PHASE IN A POWDER MIXTURE OF TWO PHASES, F. Froscher. *Journal*, Franklin Institute, Philadelphia, Pa., Feb. 1955, p. 107.

10. STUDY OF TUMOR CELL POPULATIONS BY MONTE CARLO METHODS, J. S. Hoffman, N. Metropolis, Verna Gardiner. *Science*, Washington, D. C., vol. 122, Sept. 9, 1955, pp. 265–68.

11. THE ADVANCED THEORY OF STATISTICS, M. G. Kendall. Griffin and Company, Ltd., London, England, 1948, p. 141.

Static Elimination on Bell System Teletypewriters

HAROLD BAVLEY
NONMEMBER AIEE

DONALD B. PERRY
MEMBER AIEE

B. S. SWEZEY
NONMEMBER AIEE

TELETYPEWRITERS which are well known as a means of handling written communications in industrial, government, and police business do not work properly when static is present on the paper. The use of ionizing radiation from radium salt devices to eliminate the effect of static on paper is not new. Such a device is now being used on a nationwide basis by the Bell Telephone Companies during the operation of the page and tape teletypewriters. These teletypewriters may be used in many diverse locations on the customer's premises, thus raising the question as to potential health hazards to the customer's attendants, as well as to the telephone company's maintenance forces.

The Problem

In dry winter weather or in some air-conditioned offices where the humidity is relatively low, static electicity is produced on the rolled paper in its path from the teletypewriter platen. The static may cause paper to cling to the back of the teletypewriter cabinet or adjacent wall or it may accumulate inside the teletypewriter cover. The static is caused principally by the action of the type bars in striking the paper and forcing it against the platen, and probably imbedding it momentarily into the platen as each letter is typed.

The production of static charges by the friction of one material against another can be demonstrated by rubbing a glass rod with silk to obtain a positive charge on the rod and a negative charge on the silk. It is not so widely known that static is produced merely by the contact of dissimilar materials with each other, the more nearly perfect the contact, the greater the charge. Compression of an amorphous material (such as rubber) will produce a charge in the material. Collision of one material with another if the relative velocity of motion is over a critical value will also produce a static charge. This charge appears to bear no relation to the amount of energy expended in order to cause the collision. In the case of collision, however, there is a dual effect, one charge due to contact and the other to collision. The two effects may add for some materials or be opposed for others.

Coehn[1–3] determined that the charge Q, which developed as a result of the intimate contact of materials, bore no relation to the length of time the materials were in contact but was proportional to the difference of the dielectric constants of the materials, i.e., $Q=C(K_1-K_2)$ where K_1 and K_2 = the dielectric constants. Richards[4] records his work to determine the value of the constant C and gives an expression $Q=4.43(K_1-K_2)$ where Q is coulombs per centimeter^2 of contact surface in electrostatic units assuming the dielectric constant of steel to be 3.1, and not infinity. The static charge due to contact is appreciable and the effect can be very annoying.

Paper has been known to cling to and climb a wall to the ceiling as it issued from the teletypewriter. Of course, this condition could happen only when the copy is not currently removed from the machine, as might be the case with unattended service or where the machine is automatically started before the office opens and copy is sent to that station from a station in another time zone. Also, copy is sometimes allowed to accumulate on a press wire. To avoid static difficulties, several methods have been tried. One arrangement used was to stretch tinsel with a copper center across the teletypewriter where the paper leaves the platen. The grounded tinsel in contact with the paper removes some of the static charge. However, field reports indicated that trouble still exists and this device was not put into general use.

A wire grille to contact the paper in its path from the platen was tried with unsatisfactory results. Other methods which have been considered, and some of them found insufficient, included moistening devices, rails on the cover, conducting rubber platen, conducting surface on the paper, and cork pressure rollers.

For some years the textile and printing industries have used a commercially available radioactive device to reduce the effect of static charges developed during process operations. This, known as the Ionotron static eliminator, is manufactured by the U. S. Radium Corporation of New York. This device, shown in Figs. 1 and 2, consists of a radium salt distributed throughout thin metal foil bonded to a heavier metallic nonradioactive backing which is then mounted in a suitable housing. The alpha rays and to a lesser extent the beta rays emitted by the radium salt ionize the air in the immediate vicinity so as to make the air conductive. The static is discharged continuously as it is formed. The Ionotron static eliminator for page machines being used for the Bell Telephone Companies' teletypewriters has an active source $7^1/_2$ inches long by 1/4 inch wide, and contains 60 micrograms of radium. These devices are provided with special brackets intended for rapid mounting on the teletypewriters. Each unit has a snap-on nickel steel shield which when removed can be easily attached to the bracket for storage. A silhouette view of location is shown in Fig. 3.

The nature of this radioactive static eliminator requires that precautions be taken in handling since, in addition to the relatively harmless alpha rays, radium emits beta and the more hazardous gamma radiations. The Bell Telephone Companies have issued printed instructions to each maintenance craftsman who may be called upon to install or service the teletypewriters using this means of static elimination. Basically these precautions are as follows:

1. The protective shield should be placed over the active face of the unit and removed only after the eliminator is installed in the teletypewriter.

2. The protective shield should be placed over the active face of the Ionotron when repairs or adjustments are being made to the teletypewriter.

3. When not installed on a teletypewriter, the shielded eliminator should be stored in the original shipping container provided by the manufacturer and then stored in a metal locker away from the immediate area regularly occupied by personnel.

4. In carrying the eliminator to a customer's office for installation, the shielded

---

Paper 56-106, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing November 25, 1955.

Harold Bavley is with the Massachusetts Department of Labor and Industries, Boston, Mass.; Donald B. Perry is with the American Telephone and Telegraph Company, New York, N. Y., and B. S. Swezey is with the Bell Telephone Laboratories, Inc., New York, N. Y.

Acknowledgment is made to the engineers of the Bell Telephone System, to E. Comproni, Engineer for the Massachusetts Division of Occupational Hygiene, and to Dr. Edward Shapiro and J. Conway, Health Physicist for Tracerlab, Inc., for participating in the survey.

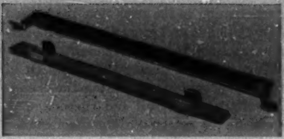


Fig. 1. Ionotron with guard removed

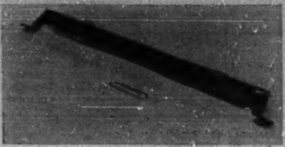


Fig. 2. Ionotron with guard in place—clip shows proportional size

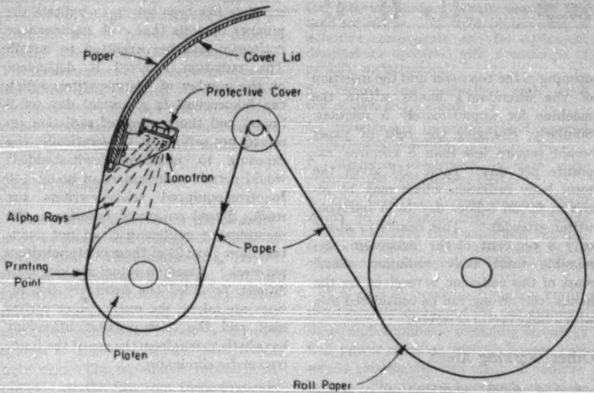


Fig. 3. Relative positions of elements in model 28 teletypewriter

Table I. Milliroentgen Dosage Received During Installation of Static Eliminator

T-35 Static Eliminator Installed in Model 28 Teletypewriter

	Dosage Rate, Mr per Hour	Exposure Time, Hours	Dosage per Installation, Milliroentgens
Direct contact with *T-35* unit			
Cover on	12	0.033	0.40
Cover off	80	0.0042	0.33
			0.73

Maximum permissible weekly exposure, 300 mr.[3]

Maximum permissible installations per day in so far as radiation exposure is concerned, 82.

Table II. Milliroentgen Dosages Received by Maintenance Men

	Dosage Rate, Mr per Hour	Exposure Time, Hours	Dosage per Servicing, Milliroentgens
Placing cover on static eliminators*	80	0.0042	0.33
Maximum possible exposure to hands, 6 inches from static eliminator	1.5	1	1.5
Removing cover from static eliminator	80	0.0042	0.33
Total dosage per instrument servicing			2.2

* Maintenance man is required to put protective cover on static eliminator before beginning maintenance work.

unit should be placed in a tool kit and not in a pocket of clothing.

5. The static eliminator should not be dusted or cleaned in the field.

6. Each craftsman is limited to 50 installations a day.

Tests to Determine Degree of Hazard

Tests were made to convince all concerned that if properly used and handled, radiation from the Ionotron static eliminator could be easily controlled.

Health physicists from Tracerlab, Inc., Boston, Mass., and engineers from the Massachusetts Division of Occupational Hygiene co-operated with the Bell System engineers in conducting these tests in which both an ionization chamber instrument and a Geiger tube survey meter were used. Working conditions involving installation and maintenance by company craftsmen in the operation of the teletypewriter, including the changing of the ribbon and paper by the customer's attendant, were simulated, using a model *28* teletypewriter.

Exposure to Maintenance Men

It was estimated that the installation craftsmen's hands would not be at any time in contact with the Ionotron static eliminator for a period greater than two minutes with the protective shield or guard on the unit or for a period of 15 seconds with the protective shield removed. This was borne out by a time study involving two individuals who performed the operation for the first time and who were not accustomed to the machine or the tools. Both of these individuals installed the unit in the teletypewriter and went through the two operations in less than 2 minutes, and in 15 seconds, respectively.

The measurements of the radiation and the estimated dosage per installation are given in Table I.

The figure of 300 milliroentgens (mr), reference 3, refers to whole body radiation.

For partial body radiation to the hands and forearms, the maximum permissible weekly dose is 1,500 milliroentgens (mr).[4]

Reference to Table I indicates that the maintenance craftsmen receive about 0.73 mr per each installation of a static eliminator in the model *28* teletypewriter. Based on the maximum allowable daily dosage, the actual dosage received by the maintenance craftsmen is almost negligible. An additional safety factor is introduced when it is realized that the maximum permissible dosage refers to total body radiation, whereas the exposure would be mainly to the craftsmen's hands during the installation of the unit. However, based on the maximum permissible total body dosage of 60 mr per day, based on a 5-day week, each installation craftsman could install 82 of the *T-35* Ionotron static eliminators without being subjected to an excessive quantity of radiation.

Maintenance

Approximately once a month it may be necessary for a maintenance man to service the teletypewriter for about 1 hour. Table II estimates the dosage of radiation which would be received by a craftsman

Table III. Milliroentgen Dosages Received by Teletypewriter Operators

	Dosage Rate, Mr per Hour	Exposure Time, Hours	Dosage per 8-Hour Day, Milliroentgens
Routine operations			
At keyboard	0.60		
Pelvic region	0.13		
Head region	0.18		
Chest region	0.23		
Knee region	0.13		
Total body maximum possible dosage	0.60	8	4.8
Special operations:			
Changing paper roll	1	0.08	0.08
changing ribbon, 5 minutes per week	1	0.017	0.02
Maximum possible dosage			4.9
Maximum permissible daily dosage			60

Table IV. General Surveys

	Dosage Rate, Mr per Hour	Total Mr per 8-Hour Day	Maximum Permissible Dosage, Per Cent
Adjacent to right side of teletypewriter	0.3	2.4	4.0
Adjacent to left side of teletypewriter	0.4	3.2	5.3
Adjacent to back of teletypewriter	0.4	3.2	5.3
Top at glass surface	15	120	200
Knee hole	0.1	0.8	1.3
Keyboard	0.6	4.8	8.0

based upon 1 hour of service time. Again this is quite small, and it is apparent that a maintenance craftsman could actually be servicing the teletypewriters as a continuous operation based on an 8-hour day, and receive about 16 mr during the entire day. The hands and forearms only are exposed to this radiation. Comparing this dosage with the maximum permissible dosage of 60 mr a day, total body radiation, it is apparent that any maintenance craftsman who follows the specified safety precautions and who would service only a few teletypewriters a day would receive a minimum amount of radiation.

Exposure to Customer's Attendant

During the operation of the teletypewriter, the attendant is seated facing the instrument so that the pelvic region, head region, and chest region are approximately 16 inches, 20 inches, and 16 inches respectively from the static eliminator. The Ionotron static eliminator is located within the machine and is further shielded by the teletypewriter parts, and the metal over the machine. During the sending operation, the attendant's hands are on the keyboard so that only the hands would be directly exposed.

The teletypewriter ribbon is changed about once a week; this requires the opening of the top cover and the insertion of the attendant's hands within the machine for approximately 5 minutes. Similarly, changing the rolls of paper would require less than 5 minutes' exposure per day. Table III gives the total body radiation which may be received for an 8-hour day by a teletypewriter attendant. This results in about only 8 per cent of the maximum permissible total body radiation. Since most of this radiation is received by the hands, this dosage can be considered safe with a large factor of safety.

Other Working Areas

The results of a radiation survey in the immediate vicinity of the no. *28* teletypewriter containing a static eliminator is shown in Table IV, which shows that the radiation levels are low except at the top glass surface. Contact with the glass surface for longer than 4 hours a day every day in the week would lead to obtaining a dosage of 300 mr for that part of the body in contact with glass. It is very unlikely that the top glass surface would be contacted either by the teletypewriter attendant or by others for this period. Other individuals working immediately adjacent to this teletypewriter would be subjected to only a small fraction of the maximum permissible dosage. If these individuals are located more than a foot from the teletypewriter, the radiation exposure to them would be negligible.

In summary, it is the consensus of those participating in the tests that the hazard of radiation is well controlled. The overall exposure to the attendant is considered minimal during the operation of the teletypewriter and during the changing of the paper roll and ribbon. The exposure per machine to the maintenance craftsman, provided the shield is kept on the active face during the installation of the Ionotron static eliminator and during the repair and adjustment of the teletypewriter, would not be excessive. A schedule has been set up to indicate the number of units that each maintenance craftsman will be permitted to install. The radiation received is dependent on the number of teletypewriters which can be serviced in a normal day (of 8 hours) and the amount of radiation received per servicing. Theoretically, the exposure to radiation (unit-shielded) would permit maintenance of about 140 Ionotron-equipped teletypewriters per week. Based on actual experience, the maintenance craftsman could not maintain more than 25 of these teletypewriters per week. Based on actual working conditions, provided the necessary precautions are taken, the maintenance craftsman and the teletypewriter attendant have little to fear from the use of the Ionotron static eliminator.

References

1. Ueber ein Gesetz der Elektrizitatserregung, **Alfred Coehn.** *Annalen der Physik*, Leipzig, Germany, vol. 64, 1898, pp. 217–32.

2. Über die quantitative Gültigkeit des Ladungsgesetzes für Dielektrika, **Alfred Coehn, U. Raydt.** *Ibid.*, vol. 30, 1909, pp. 777–804.

3. Über die Berührungselektrizität von Gasen gegen leitende und nichtleitende Flüssigkeiten, **Alfred Coehn, Hermann Mozer.** *Ibid.*, vol. 43, 1914, pp. 1048–78.

4. The Contact Electricity of Solid Dielectric, **H. F. Richards.** *Physical Review*, New York, N. Y., vol. 22, Aug. 1923, pp. 122–33.

5. Recommendations of the International Commission on Radiological Protection and of the International Commission on Radiological Units. *Handbook no. 47*, U. S. Department of Commerce, National Bureau of Standards, Washington, D. C., p. 2. (The International Commission on Radiological Protection is one of two permanent commissions operating under the auspices of the International Congress of Radiology. This Commission was first organized in 1928 and has been continuously active ever since.)

6. *Ibid.*, p. 3.

Measuring 1,800,000-Kw Demand

J. A. MORRIS
NONMEMBER AIEE

H. C. THOMAS
MEMBER AIEE

A PAPER presented by Sporn and Marquis[1] in 1954 outlines quite specifiically practically all phases of the economic, engineering, and financing problems in connection with the supply of 1,800,000 kw of power to the Atomic Energy Commission Portsmouth Area. The measurement of the electrical quantities of this load, to satisfy the contractual requirements, created a number of additional problems, not usually encountered in the measurement of power and energy.

The Problem

The problem involved in the measurement of power and energy of a load of approximately 1,800 megawatts, lies not so much in the magnitude of the load, but rather in the accounting complexities present when this load is supplied by numerous contributors over a complex system of 132- and 330-kv transmission networks. The problem is further complicated by the fact that these networks carry not only power to the 1,800-megawatt load, but also function as carriers of transfer power to and from the interconnected contributors.

Since a number of the tie-line stations are unattended, some means was necessary whereby all power and energy measurements, for two directions of power flow could be recorded simultaneously for each clock hour. It is obvious that accurate readings could be obtained only by some automatic means which would provide a simultaneously printed record for accounting purposes.

In evaluating the numerous devices available for measuring and recording power and energy, the following factors had to be considered.

1. The contractual requirement of an over-all measurement accuracy of ±1%.
2. Initial low loads at supervoltages which would gradually increase to contract value.
3. Combinations of transmission facilities which under certain conditions could change meter loadings by about 300%.
4. The high degree of reliability required for this application, which imposes the need for simple and rapid replacement of defective components, and for equipment designed for operation with a minimum of periodic maintenance.
5. Testing and calibration with usual test devices and methods.
6. Maintenance by personnel not required to have special ability.
7. Minimum investment in spare components and replacement parts.
8. The need for demand records which would permit simple analysis for billing without requiring time-consuming manipulations on special skills, and which would eliminate any ambiguity in interpretation.

A review of available metering equipment usually employed for the measurement and automatic recording of energy and power revealed that in no one case would existing devices meet all requirements. The nearest approach would be a combination of devices furnished by several manufacturers. Such an approach was considered undesirable because of the divided responsibility for operation of the equipment and the lack of operational experience with such a combination.

A further study of the problem showed that, if certain devices which had been in use by the electric utilities of the country for approximately 30 years could be modified to meet the requirements of this particular application, most of the first-considered obstacles could be removed.

Because of the wealth of operating experience with the impulse type of power metering, it was believed that certain modifications could be made in standard devices which would then permit them to meet the requirements of this particular application. The following special features were regarded as essential if the following stipulated accuracy requirements were to be met:

1. It must be possible to record up to 1,000 impulses per demand interval. The standard device is limited to approximately 300 impulses per interval.
2. Means must be provided to store impulses during the period that the demand meter is recording and resetting to zero, in a manner to avoid positively the loss of any impulses, or the recording of an impulse in the wrong demand interval.

A co-operative study by engineers of the American Gas and Electric Service Corporation and the General Electric Company resulted in a new and in some respects unique power-metering system. With equipment designed by the General Electric Company, this system satisfactorily meets the needs of this particular application.

The system uses only existing devices and designs backed up by records of satisfactory field service over long periods, with simple, yet novel modifications to provide the high impulse rate and the impulse storage means. Fig. 1 shows its components.

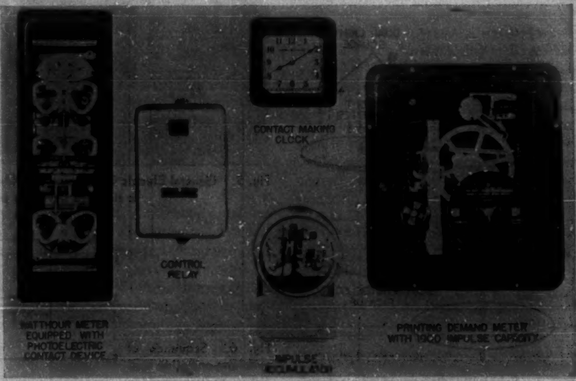
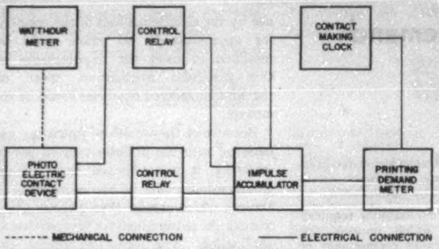

WATTHOUR METER EQUIPPED WITH PHOTOELECTRIC CONTACT DEVICE
CONTROL RELAY
CONTACT MAKING CLOCK
IMPULSE ACCUMULATOR
PRINTING DEMAND METER WITH 1000 IMPULSE CAPACITY

Fig. 1. Components of high-impulse-rate demand metering system

---

Paper 56-104, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted September 22, 1955; made available for printing November 21, 1955.

J. A. Morris is with the American Gas and Electric Service Corporation, New York, N. Y., and H. C. Thomas is with the General Electric Company, Lynn, Mass.


WATTHOUR METER
CONTROL RELAY
CONTACT MAKING CLOCK
PHOTO ELECTRIC CONTACT DEVICE
CONTROL RELAY
IMPULSE ACCUMULATOR
PRINTING DEMAND METER
------- MECHANICAL CONNECTION
——— ELECTRICAL CONNECTION

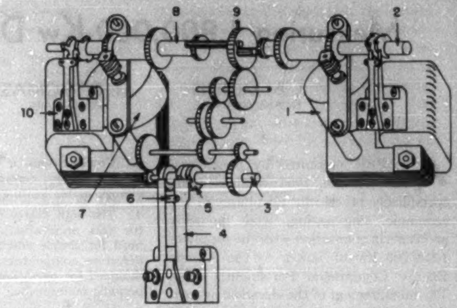

8
9
2
10
1
6
7
5
3
4

Fig. 2 (above). Block diagram of high-impulse-rate demand metering system

Fig. 4 (right). Schematic diagram of impulse accumulator

Solution

Fig. 2 is a block diagram to illustrate functionally the high impulse demand metering system. The watt-hour meter measures energy consumption in the usual manner. To meet the requirements for sustained accuracy over the entire load range, a photoelectric contact device was designed to replace the conventional mechanical contact device. A schematic diagram of this new mechanism is shown in Fig. 3. It acts as a single-pole double-throw switch, alternately moving from one position to another in response to the rotation of the watt-hour meter rotor. A control relay is connected in the output of each switch position to provide increased energy-handling capacity. Gearing within the contact device permits preselection of the energy consumption necessary to initiate switching action.

The switching action of the control relays is arranged to generate electric impulses, each impulse representing a specific value of energy consumption. These impulses are first fed to the impulse accumulator, which is used with each printing demand meter to provide for the storage of impulses generated during the period in which the demand meter is printing and resetting to zero. Its use insures against the possibility of losing impulses, and makes possible a precise check of the total impulses generated against the total recorded by the demand meter.

The impulse accumulator is shown schematically in Fig. 4. In many respects, it is mechanically similar to the General Electric Company type *MD-4* totalizer equipped with a type *D-6* contact device, but it functions differently. An incoming impulse operates motor 1 to rotate shaft 2 through 30 degrees and, through differential gears 9 and an intermediate gear train, to rotate cam shaft 3, in order to lift brush 4 out of the slot in cam 5 and close contact 6. The closing of contact 6 energizes motor 7 which rotates shaft 8 and, through differential gears 9 and the intermediate gear train, to rotate cam shaft 3, in order to return cam 5 to its "home" position, where contact 6 opens and de-energizes motor 7. The outgoing contact device 10 is also operated by shaft 8, to generate one outgoing impulse for each incoming impulse.

To store impulses, the electrical circuit between contacts 6 and motor 7 is opened by an auxiliary switch, normally located in the contact-making clock. Incoming impulses then continue to rotate cam shaft 3, each impulse pro-

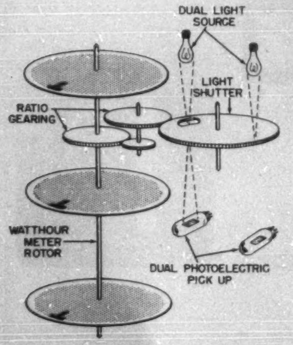

DUAL LIGHT SOURCE
LIGHT SHUTTER
RATIO GEARING
WATTHOUR METER ROTOR
DUAL PHOTOELECTRIC PICK UP

Fig. 3. Schematic diagram of photoelectric contact device

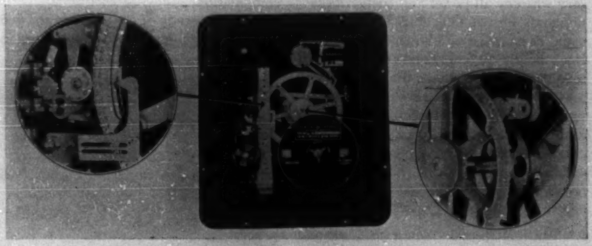


Fig. 5. General Electric Company type-PD-6Y1 printing demand meter showing modifications to standard device to provide 999-impulse capacity

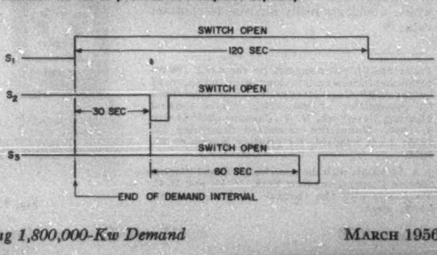

SWITCH OPEN
120 SEC
S_1
S_2
SWITCH OPEN
30 SEC
S_3
SWITCH OPEN
60 SEC
END OF DEMAND INTERVAL

Fig. 6. Sequence of switching operation special contact-making clock

ducing a definite angular rotation. When the auxiliary switch is again closed, motor 7 remains energized until cam 5 returns to its home position at which time an outgoing impulse has been generated for each stored impulse. Stored impulses are converted into outgoing impulses at the rate of one about every 2 seconds. The mechanism permits incoming impulses to be received at the same time that outgoing impulses are being generated, without any loss of impulses.

Output impulses from the impulse accumulator are fed to the printing demand meter where they are counted and recorded in the usual manner. The need for a printing demand meter capable of accepting up to 1,000 impulses during each demand interval was met by modifying the General Electric Company type *PD-6* printing demand meter. The demand registering mechanism was redesigned to accept 999 impulses by providing additional numerals on the small type wheel and modifying the mechanism which indexes this wheel. The reset mechanism was redesigned to be operated by a continuously energized motor rather than by the storage of energy in a spring. By this means the reset torque is maintained at a constant value and reset is accomplished at a relatively constant speed, with minimum acceleration and deceleration. A further modification of the reset mechanism permits the reset operation to be accomplished with the large type wheel resetting in a maximum of one revolution. Fig. 5 illustrates the design modifications which have been made.

The special contact-making clock, a modification of the standard General Electric Company type *C-16* contact-making clock, provides the necessary switching action by means of three independent switches operated in time sequence, as shown in Fig. 6. The first switch S_1 opens at the end of the demand interval to de-energize the output mechanism of the impulse accumulator and thus shuts off the flow of impulses to the printing demand meter. (This switching action, in effect, defines the end of one demand interval and the start of the next.) Thirty seconds later, the second switch S_2 closes and sends an impulse to the demand meter, causing it to print the interval demand and to reset to zero. Sixty seconds later, the third switch S_3 closes and sends another impulse to the demand meter, causing it to print "zero" and thus provide a record that the demand meter has reset to zero. Thirty seconds later the first switch S_1 closes and re-energizes the output mechanism of the impulse accumulator. During the 2-minute period in which the output mechanism of the impulse accumulator is de-energized, all incoming impulses are stored; none is lost or counted in the wrong demand interval.

Conclusion

In conclusion, it should be emphasized that this paper illustrates the beneficial results and advancements in the metering art which can be realized through the co-operative effort of a utility and a manufacturer, wherein application requirements, design and manufacturing problems, and difficult delivery schedules were recognized as elements of a single problem. In this case through such co-operative effort special equipment was successfully provided to meet a special demand-metering application, and which has pointed the way to improved techniques for demand metering in general.

Reference

1. The OVEC Project: Economic, Engineering, and Financing Problems of the 2,200,000-Kw 18,000,000,000 Kilowatt-Hour Power Project of the Ohio Valley Electric Corporation, Philip Sporn, V. M. Marquis. *AIEE Transactions*, vol. 73, pt. III-A, Feb. 1954, pp. 49–63.

A Signal-Discriminating Magnetic Amplifier

G. E. LYNN
ASSOCIATE MEMBER AIEE

J. F. RINGELMAN
ASSOCIATE MEMBER AIEE

A NEW FIELD of application for magnetic amplifiers is promised by the development of a signal amplitude discriminator. This signal-discriminating amplifier has the unique ability to select the largest of any number of signals and to amplify the selected signal in the normal way. The mechanics of the amplifier are described in this paper and typical measured values are discussed.

Voltage Summation or Selection

In view of the many applications of magnetic amplifiers as summing devices, the existence of a magnetic phenomenon which would allow signal selection rather than summation when first considered might seem improbable. When time-varying voltages are involved, however, further thought reveals that summation of these voltages on a single magnetic reactor is really the improbable result, as will be shown.

Consider, for instance, a toroidal reactor with several windings such as is commonly used in a magnetic amplifier. If the windings are assumed to have zero resistance and if leakage flux is negligible, the volts per turn must be the same for all windings. This will be true regardless of the number or comparative magnitudes of external voltage sources included in the circuits of which the individual windings form a part. If impedances external to the reactor windings in the individual circuits are small, the volts per turn ratio will be determined by the source which supplies the highest voltage for the number of turns to which that voltage is applied. All other windings will deliver energy to their respective circuits rather than accept energy from them.

For this reason, a single-core magnetic amplifier which has several control windings to which time-varying voltages are applied will be controlled by the signal supplying the highest volts per turn ratio. If loading between windings is prevented, the other signals will not affect the operation of the amplifier.

The problem of signal selection rather than summation may be resolved into two parts:

Paper 56-226, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted September 30, 1955; made available for printing December 7, 1955.

G. E. Lynn and J. F. Ringelman are with Westinghouse Electric Corporation, Baltimore, Md.

The authors wish to acknowledge the unpublished work of F. A. Baker, formerly with Westinghouse Electric Corporation, concerning the development of the method of impedance control described in the "Analysis of Basic Circuitry" portion of this paper.

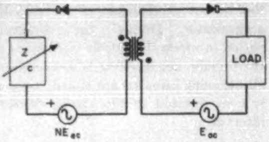


Fig. 1. Ramey-type magnetic amplifier circuit

1. Conversion of all signals to time-varying voltages of one frequency whose average value is proportional to signal amplitude.
2. Elimination of the loading between circuits when these time-varying voltages are coupled to a common magnetic reactor.

Analysis of Basic Circuitry

Ramey has described a magnetic amplifier, Fig. 1. which applies a variable volt-second integral to reset a magnetic core.[1,2] Variation of the volt-second integral is obtained by use of a variable impedance in the control circuit. Since a magnetic amplifier is itself a variable impedance, it is a simple matter to substitute the gate winding of another magnetic amplifier (of the basic half-wave type) for the control impedance of Fig. 1, as shown in Fig. 2. The load voltage E_L is now a function of the signal voltage E_s.

In explaining the operation of Fig. 1, certain simplifying assumptions will be made:

1. Cores L_1 and L_2 have flux-current loops essentially rectangular, as shown in Fig. 3.
2. Rectifiers X_1 and X_2 are ideal; i.e., zero conduction in the reverse direction and zero voltage drop in the forward direction.
3. The resistance in the windings of the reactors is negligible.
4. When applied to winding N_L, the magnitude of E_{ac} is just sufficient to drive the flux of reactor L_2 from negative to positive saturation in one half-cycle of supply frequency. Reactor L_1 is designed so that NE_{ac} will just drive flux from negative to positive saturation in one half-cycle of supply frequency.
5. The exciting current of the gate winding of reactor L_1 is negligible in comparison with the exciting current of the control winding of reactor L_2.

As a further aid to explanation, the basic impedance controlled amplifier will be broken down into three circuits:

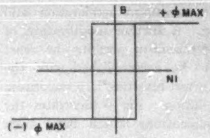


Fig. 3 (above). Idealized flux-current loop

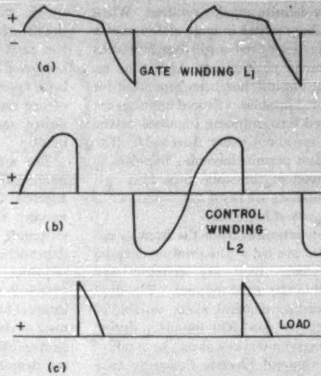


Fig. 4 (right).

a—Gate-winding voltage of reactor L_1
b—Control winding voltage of reactor L_2
c—Load voltage

1. The *signal circuit*, which consists of the series combination of:
   (*a*) Signal source E_s.
   (*b*) Total circuit resistance, R_s.
   (*c*) Control winding N_s of reactor L_1.
2. The *control circuit*, which consists of the series combination of:
   (*a*) Voltage source NE_{ac}.
   (*b*) Gate winding N_g of reactor L_1.
   (*c*) Ideal rectifier X_1.
   (*d*) Control winding N_c of reactor L_2.
3. The *load circuit*, which consists of the series combination of:
   (*a*) Voltage source E_{ac}.
   (*b*) Gate winding N_L of reactor L_2.
   (*c*) Ideal rectifier X_2.
   (*d*) Total load circuit resistance R_L.

Two distinct periods of time will be considered and will be distinguished from each other by the polarity of E_{ac}. A positive half-cycle will be defined as a half-cycle when E_{ac} tends to cause forward current through rectifier X_2. A negative half-cycle, then, is a half-cycle when NE_{ac} tends to cause forward conduction in rectifier X_1.

The reactor windings are shown dotted on one end to show relative instantaneous voltage polarities. The dotted ends of all windings on one core have the same instantaneous polarity.

In analysis, let it first be assumed that E_s is zero and the control circuit is open. Voltage E_{ac} will cause reactor L_2 to saturate on the first positive half-cycle of supply voltage. Each succeeding positive half-cycle will then appear across load resistor R_L, causing load voltage E_L to have its maximum value. If the control circuit is now closed and E_s is maintained at zero, reactor L_1 will saturate in the first negative half-cycle of supply voltage. Since it was assumed that the exciting current of the gate winding of reactor L_1 is very small compared with the exciting current of the control winding of reactor L_2, negligible voltage will appear across the control winding of reactor L_2 during a negative half-cycle until reactor L_1 saturates. Following saturation of reactor L_1, each succeeding negative half-cycle of supply voltage will cause NE_{ac} to appear across the control winding of reactor L_2. If on the positive half-cycles of supply voltage, reactor L_2 is driven from $-\phi_{max}$ to $+\phi_{max}$ by E_{ac} on the negative half-cycles it will be driven from $+\phi_{max}$ to $-\phi_{max}$ by NE_{ac}. Under these conditions, E_{ac} will appear across the gate winding for each entire positive half-cycle and no voltage will appear across R_L. This latter statement is, of course, only approximately true, since the exciting current of the gate winding of reactor L_2 will flow through R_L. Where the resistance of R_L is very small with respect to the unsaturated impedance of the gate winding, the approximation is a valid one.

It has now been shown that, by application of voltage NE_{ac} in the circuit of Fig.

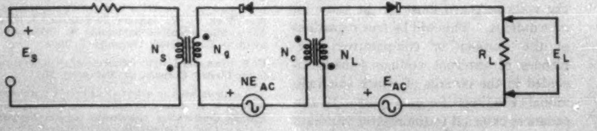


Fig. 2. Basic signal converter used with Ramey-type amplifier

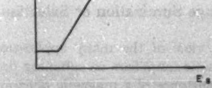


Fig. 5. Average output voltage as a function of average signal volts for circuit of Fig. 2

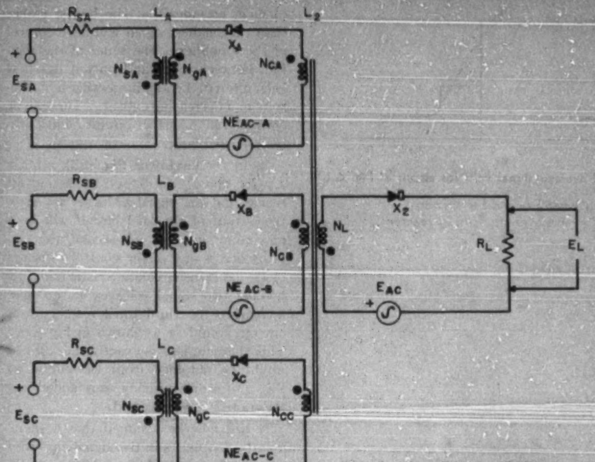


Fig. 6a. Basic signal-discriminating amplifier circuit

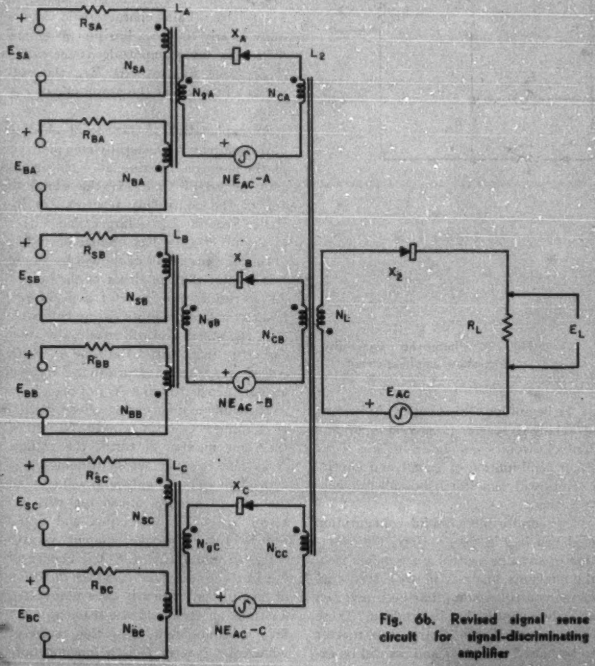


Fig. 6b. Revised signal sense circuit for signal-discriminating amplifier

2, the load voltage E_L has been changed from a maximum to essentially zero. The next step is to examine the possibilities of obtaining intermediate voltages. If a small signal voltage E_s is applied, the flux in reactor L_1 will be driven from saturation to some unsaturated condition during the positive half-cycles of supply voltage. On the negative half-cycles, NE_{ac} will appear across the gate winding of reactor L_1 until the core has absorbed sufficient volt-seconds to bring the flux back to saturation. (Integration of both sides of the expression $e=Nd\phi/dt$ shows that $(\phi_2-\phi_1)=1/N\int edt$. For a given winding, then, a change of flux in a reactor is equivalent to a volt-second area.) After saturation, NE_{ac} will appear across the control winding of reactor L_2 for the remainder of the half-cycle. Since some of the volt-second area represented by a half-cycle of NE_{ac} has already been expended in changing flux in reactor L_1, a sufficient area will not be available to change flux in reactor L_2 all the way from positive saturation to negative saturation. At the end of the negative half-cycle, therefore, the flux in reactor L_2 will be somewhere between positive and negative saturation. As a result, only a portion of the positive half-cycle volt-second area supplied by E_{ac} will be required to drive the flux in reactor L_2 back up to positive saturation. The remainder will appear across R_L after saturation is reached.

The wave shapes appearing across particular components for a typical condition are shown in Fig. 4. In Figure 4(a), the voltage appearing across the gate winding of reactor L_1 during the positive half-cycle results from the flux change during reset. The shape of this voltage is a function of the impedance and net voltage of the control circuit as well as the magnitude and shape of the signal voltage and the signal circuit impedance. The wave shape shown is for a small d-c signal and a moderately high signal circuit impedance. Fig. 4(b) indicates that NE_{ac} appears across the terminals of the control winding N_c for some portion of the negative half-cycle and that an equal voltage is induced on the control winding during the positive half-cycle by application of E_{ac}. Fig. 4(c) illustrates the wave shape across the load resistor R_L.

If the signal voltage E_s is sufficiently large, the flux in reactor L_1 will be driven to negative saturation during its reset half-cycle. NE_{ac}, then, will appear across reactor L_1 for the entire negative half-cycle and no voltage will appear across the control winding of reactor L_2

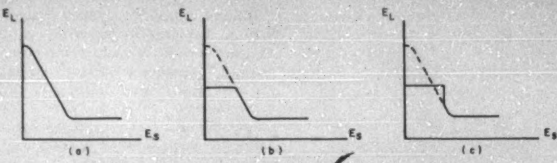


Fig. 7. Average output voltage as a function of average signal volts for circuit of Fig. 6(b)

a—Characteristic obtained with reactor L_A when E_{SB} and E_{SC} are zero
b—Characteristic obtained when reactor L_B takes over control from reactor L_A
c—Discontinuity due to coupling effects

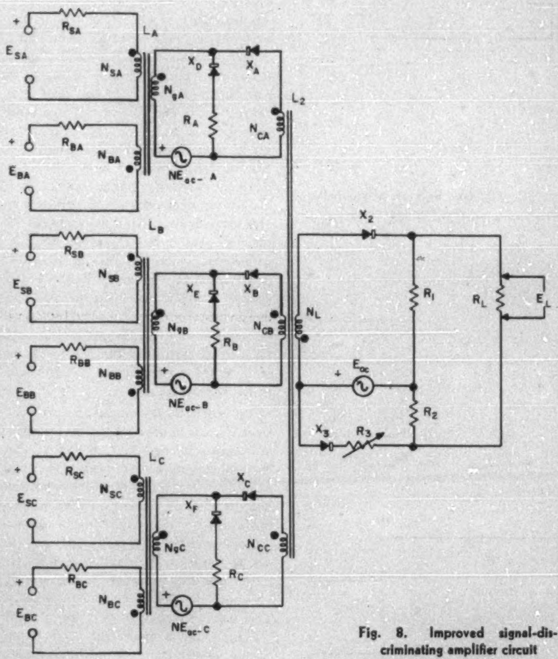


Fig. 8. Improved signal-discriminating amplifier circuit

during that time. Under these conditions, E_{ac} will appear across the load for the entire positive half-cycle. Thus, the signal voltage can control the load voltage from zero to maximum, as shown on the transfer characterictic of Fig. 5.

The Signal-Discriminating Amplifier

It has now been shown that, by the use of reactor L_1, a signal can be converted into a time-varying voltage. The first half of the 2-part problem mentioned at the beginning of this paper is solved by attaching to reactor L_2 several signal and control circuits, as shown in Fig. 6(a). Each combination of signal and control circuits will henceforth be called a control stage.

A generally more useful configuration is shown in Fig. 6(b). Here, the operation of any one control stage differs from the previous analysis in that the sense of the control winding has been reversed and a bias winding has been added. These changes permit each control stage reactor to be biased to cut off and, as will be explained, provides that the largest rather than the smallest signal will have control of the amplifier. The sense of the relation between input and output has now been altered; i.e., an increasing signal results in a decreasing output. This can be corrected by a further circuit addition to be discussed later.

If, in the circuit of Fig. 6(b), reactor L_A, L_B, and L_C are biased to the cut off region and the signal voltages are zero, the output of the load circuit will be a maximum since no volt-second area appears across any of the control windings of L_2. A transfer characteristic taken by holding two signal voltages at zero and increasing the third until reactor L_2 is cut off, would be as shown in Fig. 7(a). This characteristic is exactly the same as would be obtained from the circuit of Fig. 2, if a bias winding were added and the control sense reversed.

It will now be assumed that a signal voltage E_{SA} exists in the signal circuit of reactor L_A of sufficient magnitude to cause an output voltage E_L, of about one half the maximum shown in Fig. 7(a). Increasing the signal voltage E_{SB} of reactor L_B from zero to maximum results in the transfer characteristic shown in Fig. 7(b). This characteristic shows quite clearly that E_{SB} has no effect until it has the same amplitude as the signal E_{SA}. Once E_{SB} exceeds E_{SA}, the output is a function of the amplitude of E_{SB} only.

An explanation of this signal-selecting operation requires a step-by-step analysis of the voltages existing at each significant period of time. If the signal to one of the controlling reactors, for instance reactor L_A, is larger than either of the others, the flux change occurring during the reset half-cycle will be less in this reactor than in either of the others. As a result, this reactor will absorb fewer volt-seconds in its gating half-cycle than the others and will reach saturation earlier in time. After the time at which reactor L_A reaches saturation, $NE_{ac}-A$ appears across winding N_{CA} of reactor L_2 in the same manner as described earlier. If windings N_{CA}, N_{CB}, and N_{CC} have the same number of turns, the voltage NE_{ac} also appears across windings N_{CB} and N_{CC}. These induced voltages in the control circuits associated with reactors L_B and L_C are equal and opposite to the respective source voltages $NE_{ac}-B$ and $NE_{ac}-C$. The change of flux in reactors L_B and L_C in the direction of positive saturation is thus terminated for this half-cycle and the reactors never fire. It is apparent, then, that the reset voltage for reactor L_2 is a function only

Vol. 7
Pt. 1
1956

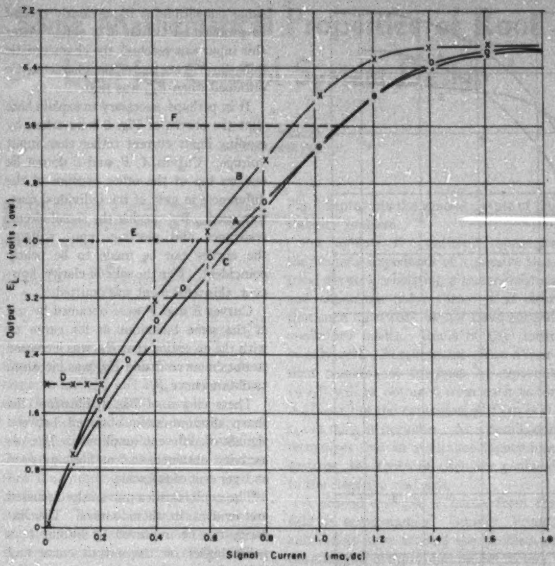


Fig. 9. Individual and combined transfer characteristics of the signal-discriminating amplifier shown in Fig. 8

of the control circuit associated with reactor L_A and, hence, the output voltage E_L is a function only of the signal E_{SA}.

It will be assumed, for the moment, that the flux levels existing in reactors L_B and L_C at the start of a negative half-cycle are a function exclusively of the signal voltages E_{SB} and E_{SC}. With this assumption in mind, it can be seen that, as some other signal voltage, for instance that of reactor L_B, is increased, the theoretical firing angle will decrease in the associated gate circuit until it is finally less than the actual firing angle of reactor L_A. When the theoretical firing angle is less than that of the reactor which had been in control, reactor L_B will gain control of L_4 and decrease the output voltage as shown in Fig. 7(b).

EFFECTS OF NONIDEAL CONDITIONS

There exists one most significant departure in the actual operating circuit from the simplifying assumptions stated before. In reality, the flux reset levels of reactors L_A, L_B, and L_C are dependent not only on their respective bias and control currents but upon the difference in their firing times. If it is assumed, for instance, that the signal to reactor L_A is slightly greater than the signals to reactors L_B and L_C, then, after firing of L_A, a voltage equal to $NE_{ac}-A$ will appear across control winding N_{CB} as previously discussed. (The effects on reactor L_C will be the same as for reactor L_B. For the sake of brevity, only reactor L_B will be discussed.) With the appearance of this voltage across N_{CB}, flux change in reactor L_B in the positive direction will cease. This follows from the fact that voltage $NE_{ac}-B$ and the voltage induced on N_{CB} are equal and opposite, resulting in a net voltage of zero.

At the moment that the net voltage becomes zero, flux reset begins in reactor L_B under the net influence of the bias and signal voltages. It is important to note that this reset action starts during a gating half-cycle. Flux reset does not occur instantaneously, but requires a definite length of time to proceed to completion. The time required is a function of the circuitry involved. Under the conditions imposed by the circuit of Fig. 6, reset which starts at the beginning of a reset half-cycle is interrupted by the start of the next gating half-cycle. This can be seen in the wave shape of Fig. 4(a). This indicates that, given more time, a larger flux change would have been achieved. When reset starts in the gating half-cycle, as in the case under discussion, more time is allowed for reset and a larger flux change actually is achieved.

An interaction now exists between reactors L_A and L_B. If reactor L_A had not fired, the gating half-cycle of reactor L_B would not have been interrupted and the flux level in reactor L_B would have been higher at the start of the next gating half-cycle.

For a given signal amplitude, then, at least two flux levels are possible in reactor L_B at the start of its gating half-cycle. One is obtained when reactor L_A does not fire first, the other when reactor L_A has fired first. When the signal E_{SB} is increased until reactor L_B takes over control, there is an abrupt transition from the latter condition to the former. This transition results in a discontinuity in the amplifier transfer characteristics similar to that shown in Fig. 7(c).

The addition of a rectifier and a resistor to each control circuit, as shown in Fig. 8, overcomes this deficiency. Now, if reactor L_A is in control, when reactor L_A fires a voltage equal to $NE_{ac}-A$ will appear across control windings N_{bB} and N_{CC} as before. In the control circuit associated with reactor L_B, this voltage will no longer prevent current flow in the gate winding N_{gB}. The voltage rise around the circuit formed by $NE_{ac}-B$, winding N_{gB}, X_B, and R_B remains $NE_{ac}-B$. Reactor L_B is now allowed to saturate in its gating half-cycle and the flux level existing at the start of a gating half-cycle is a function only of signal E_{SB}. The dependence upon the amplitudes of the other signals has been eliminated.

In the foregoing analysis, the effects of reverse leakage and forward drop in the diodes have been ignored. By using silicon diodes it is possible to obtain leakage currents so low that they properly may be dismissed. The problem of forward drop, however, remains. As a result of this drop, the voltage appearing across the control windings of reactor L_1 after one of the controlling reactors has fired is not NE_{ac}, but is smaller by the amount of instantaneous voltage appearing across diode X_A, X_B, or X_C. In designing the reactors for the circuit of Fig. 8, the effect of forward drop is considered only when calculating the number of turns to be wound on reactor L_2. In the circuit of Fig. 6, however, the effect of forward drop of the controlling circuit diode may be to allow flux change in reactors L_B and L_C to continue after reactor L_A has fired, although at a much slower rate. For this reason, the magnitude of the discontinuity encountered with this circuit may vary.

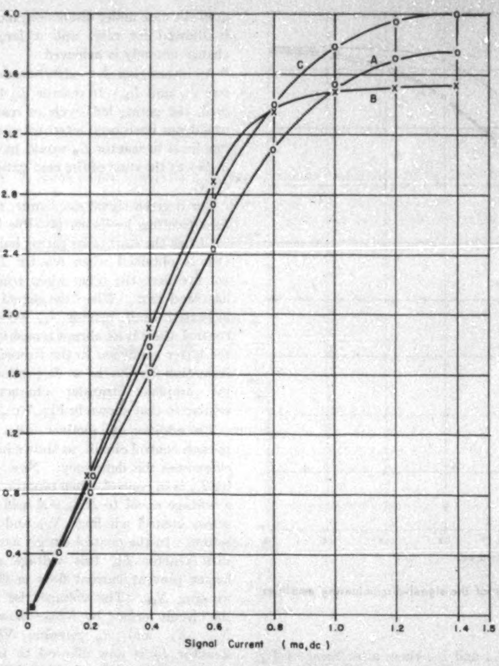


Fig. 10. Individual transfer characteristics of the signal-discriminating amplifier with bias point shifted to improve linearity

A departure from ideal conditions which is less amenable to solution is a nonlinearity in the transfer characteristic inherent in the circuit of Fig. 2. It apparently arises from the effect, on reset of the controlling reactors, of the variation of the net voltage rise around their gating circuits during their reset half-cycle. The effect, on an amplifier's reset characteristics, of the voltage applied to the gate winding of a half-wave magnetic amplifier during the reset half-cycle is discussed by Roberts, Huhta, and others.[3,4] In the circuit of Fig. 2, this effect is complicated by the fact that the voltage seen by the reactor is not constant but is a function of the firing angle of reactor L_2. Attempts to remove or decrease this effect by the use of switching transistors to replace diodes X_A, X_B, and X_C have not been entirely successful at the time of writing.[5] The nonlinearities can be removed from the circuit of Fig. 2 by this method, but when applied to the circuit of Fig. 8, a small interaction between signal voltages becomes apparent in the transfer characteristics (not shown).

EXPERIMENTAL RESULTS

Transfer characteristics obtained from the signal-discriminating magnetic amplifier circuit of Fig. 8 are shown in Fig. 9. To reverse the characteristic of Fig. 7 and allow proper sensing of input and output (zero output for zero signal), resistors R_1, R_2, and R_3, and diode X_3 have been added to the load circuit, as shown.

The curves of Fig. 9 were obtained in the following manner. With all signal voltages zero at the start, each signal in turn was increased to a maximum in small increments and returned to zero, while readings of input and output were taken at each step in the usual manner. These readings are plotted as curves *A*, *B*, and *C*, corresponding to signal voltages E_{SA}, E_{SB}, and E_{SC}, respectively. Load voltage E_L was then increased to 2 volts by increasing E_{SA} to the proper magnitude. Voltage E_{SB} was then increased in small increments until an output voltage of 4 volts was reached. These points are plotted as curve *D*. Note that the output was constant until E_{SB} reached the same magnitude as it previously needed to produce an output of 2 volts. After this input was reached, the characteristic falls directly on top of the characteristic obtained when E_{SA} was zero.

It is, perhaps, necessary to explain here that the curves of Fig. 9 were taken by reading input current rather than input voltage. Curves *A*, *B*, and *C* do not lie one on top of the other because of the differences in gain of the individual reactors used. By varying the signal circuit resistance and plotting input voltage, the curves can be made to be nearly coincident. For the sake of clarity, however, this refinement was omitted.

Curves *E* and *F* were obtained by use of the same technique as for curve *D*, with the exception that E_{SC} was increased to obtain curve *E* and E_{SA} was increased to obtain curve *F*.

These curves of Fig. 9 illustrate the sharp discrimination obtained between signals of different amplitude. The selectivity obtained was at least as good as 1 per cent of full scale.

The nonlinearities previously discussed are evident in these curves. The linearity can be improved by biasing to a point higher on the output curve and balancing out the resulting output voltage by changing the value of resistor R_3. Fig. 10 shows curves obtained by use of this method, yielding a linearity of about 2 per cent of full scale.

Conclusion

A circuit capable of discriminating between signals of different amplitudes has been presented. The circuit is not limited to either a-c or d-c signals, but operates equally well with either or both. It responds to the average value of a signal and will compare this value with the average of some other signal. A circuit of this type can readily find application in protection devices, telemetering, and other fields.

References

1. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, Robert A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

2. ON THE CONTROL OF MAGNETIC AMPLIFIERS. *Ibid.*, pp. 2124–28.

3. MAGNETIC CHARACTERISTICS PERTINENT TO THE OPERATION OF CORES IN SELF-SATURATING MAGNETIC AMPLIFIERS, R. W. Roberts. *Ibid.*, vol. 74, pt. I, Jan. 1955, pp. 682–90.

4. FLUX RESETTING CHARACTERISTICS OF SEVERAL MAGNETIC MATERIALS, Hoobert Huhta. *Ibid.*, vol. 73, pt. I, May 1954, pp. 111–14.

5. TRANSISTOR SWITCHES IMPROVE MAGNETIC AMPLIFIER PERFORMANCE, R. L. Bright, R. O. Decker, R. A. Ramey. *Electrical Manufacturing*, New York, N. Y., Apr. 1955, pp. 98–101.

Some Mathematical Properties of Root Loci for Control System Design

F. M. REZA
NONMEMBER AIEE

IN the study of a linear lumped system one is confronted with a linear differential equation

$$a_0i+a_1\frac{di}{dt}+\ldots a_n\frac{d^ni}{dt^n}=b_0e+b_1\frac{de}{dt}+\ldots b_m\frac{d^me}{dt^m} \quad (1)$$

where the applied stimulus and system response are $e(t)$ and $i(t)$, or vice versa, and t is the independent variable.

When one is dealing with physical systems, either the physical realizability or the stability of the system requires that the transient response of a system with loss vanish for a vanishing applied stimulus. This characterization somewhat restricts the location of the roots of the characteristic equation of the differential equation of the system. Indeed, if one uses operational methods, the transient and steady-state solutions of the system, equation 1, will be described by equation 2

$$\frac{E}{I}=\frac{P(s)}{Q(s)}=\frac{a_0+a_1s+\ldots a_ns^n}{b_0+b_1s+\ldots b_ms^m} \quad (2)$$

where $s = \sigma + j\omega$ is a complex variable. A general restriction on the system could be expressed conveniently as a restriction on the location of roots of the polynomials $P(s)$ and $Q(s)$.

The first task in this work is to realize the fact that the general problem of the stability of linear systems depends upon the study of a fundamental problem of mathematics, namely, knowledge of the approximate location of the roots of a polynomial in a region of the complex plane, and the loci of these roots as functions of the coefficients of that polynomial. Fig. 1 represents the general nature of the stability problem.

For instance, if equation 2 represents the transfer impedance of a passive electrical network consisting of linear resistors and capacitors only, then it is clear that such a network cannot have natural oscillatory modes. Roots of $Q(s)$ must be only real. Furthermore, since a transient response of the form A_k exponent $(+s_kt)$ will be obtained from each factor $A_k/(s-s_k)$ of the expansion of $P(s)/Q(s)$, s_k can only be negative. As a result, one recognizes that in resistance-capacitance systems, the roots of $Q(s)$ are confined to the negative real axis.

A second example is taken from the field of aerodynamics. Gunder, Friant, and others have studied the problem of the chugging of liquid-fuel rocket motors.[1] A rocket motor is stable (chug-free) if, and only if, none of the roots of

$$s^2+Ms+N+\epsilon^{-s}(Ls+T)=0 \quad (3)$$

are in the right-half s-plane (with some restriction for roots on the axis of imaginaries); M, N, L, and T are numerical coefficients depending on the system under consideration.[2] In the study of such problems, one is interested to find the loci of the roots of this equation while one or more of the coefficients is varying.

A third example is taken from a more recent field, namely, the linear programming of digital computers. The transfer function of a linear real-time program is generally expressed as

$$W(s)=\frac{a_0+a_1\epsilon^{-Ts}+a_2\epsilon^{-2Ts}+\ldots+a_n\epsilon^{-nTs}}{b_0+b_1\epsilon^{-Ts}+b_2\epsilon^{-2Ts}+\ldots+b_m\epsilon^{-mTs}} \quad (4)$$

A transformation of the exponential type

$$z=\epsilon^{-Ts} \quad (5)$$

leads back to the familiar field of the rational fractions

$$W(z)=\frac{a_0+a_1z+a_2z^2+\ldots+a_nz^n}{b_0+b_1z+b_2z^2+\ldots+b_mz^m} \quad (6)$$

The stability of the system requires that all the roots of the denominator of equation 6 remain outside the unit circle of the z-plane. Roots on the unit circle may not be multiple.[5,6]

Although a number of classical tests are available to determine whether a polynomial has all its roots within the unit circle, still a deeper insight into the problem is desirable. It is of greater interest to obtain a priori knowledge of the whereabouts of the roots without being engaged in the numerical solution of these equations. Having this in mind one is naturally led to the study of certain potential curves which curves are the topic of this paper.

From the previous examples, it is concluded that a discussion of the stability of a lumped linear system can be translated mathematically into the following problem: Study of the relative location of the roots of a polynomial in a region of the complex plane.

The first objective in this section of the paper is to unify the aforementioned two well-known problems of stability and root locus. Indeed, it is unfortunate that study of the polynomials and rational fractions such as referred to in the text and the comprehensive bibliographies of references 5 and 6 has rarely been made by engineers dealing with control problems. It is also essential to estimate the complexity of an engineering problem in this field by first translating it into mathematical terms where more tools are available.

In most engineering problems of this kind, the engineer's work is not fully done by simply guaranteeing the confinement of the roots of a polynomial to a domain as a simple application of Routh or similar criteria. A physical system usually contains a number of parameters; moreover, the characteristics of the involved elements such as vacuum tubes, springs, etc., vary with time. Some knowledge of the relative position of the roots in the frequency domain is extremely valuable. In this respect, some engineering methods devised for this problem in the last decade form an addition to the well-known earlier classical work on stability. As an example, consider the stability of a system with a feedback loop; see Fig. 2. The stability of the system requires that all poles of $W(s)$ remain in the left-half s-plane.

$$W(s)=\frac{k_1Z(s)}{1+k_1k_2Z(s)}=\frac{k_1P(s)}{Q(s)+k_1k_2P(s)} \quad (7)$$

where $Z(s)=P(s)/Q(s)$. As k_1k_2 is varied one may study the loci of the roots of

$$Q(s)+k_1k_2P(s)=0 \quad (8)$$

These are the curves which may be called "root loci." Evans, who is largely responsible for the use of root locus in

---

Paper 56-125, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 24, 1955; made available for printing December 14, 1955.

F. M. Reza is with Syracuse University, Syracuse, N. Y.

The author wishes to thank Prof. J. L. Walsh of the Mathematics Department of Harvard University, who is a well-known authority on potential curves and related topics, for his stimulating discussion and advice; Dr. W. R. LePage of Syracuse University, and Dr. W. W. Seifert of the Dynamic Analysis and Control Laboratory at Massachusetts Institute of Technology for reading the manuscript.

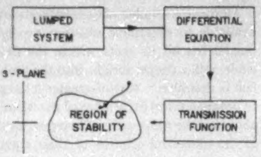


Fig. 1. General nature of the stability problem

servomechanisms, suggests the merit of the study of the root locus in control problems. He gives a number of very interesting examples of application; however, he does not describe or discuss the geometry of these curves.[7–] In the second section for the convenience of the reader is given a summary of the practical methods of finding root loci. Fig. 3 represents an example of the root-locus problem.

The second objective, in the third section of the paper, is the study of the generalized root locus curves pertaining to a polynomial $P(s)=k$, when k is a variable parameter. $Q(s)$ and $P(s)$ could be polynomials with real or complex coefficients; k may also be a complex parameter. At this stage it is unimportant that such systems are or are not of immediate engineering value. The aim is to attack a problem of mathematical physics in its utmost generality.

The fourth section is similar to the third with the exception that it deals with root locus pertaining to rational fractions.

In the fifth section is presented a unified picture of the root locus and well-known potential curves. Most of the mathematical developments of the topic were already known before the end of the last century, but apparently they have not been fully exploited for engineering purposes. In fact, the engineering literature pertaining to the subject seldom mentions this mathematical connection, and most of the properties already known to the mathematical physicists of the 19th century, such as Briot, Bouquet, Gauss, Klein, Lamé, and Maxwell, are being rediscovered for their engineering use. Undoubtedly the recent engineering literature of the subject is of great assistance in the design of servo control systems. Nonetheless it is felt that the connection of these studies with our classical background has not been fully discussed elsewhere. In this respect it is hoped that this paper will revive some interest in classical sources as a basis for further exploration and application of control problems.

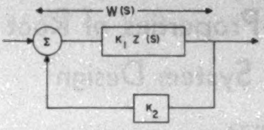


Fig. 2. General single-loop feedback system

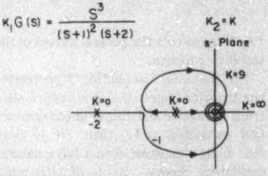


Fig. 3. Example of root-locus problem

Summary of Root-Locus Methods

At present the root-locus methods deal only with the case of a sinlge-loop feedback with real gain factors and real transfer functions using the common notation.

$$1+kG(s)=0 \tag{9}$$

Root locus is defined as the locus of roots at equation 9 when k varies from zero to infinity. Bollay[10] and Yeh[11] summarize the properties of the root-locus curve in the following:

1. A branch of root locus starts from an open-loop pole where $k=0$ and terminates at an open-loop zero where $k=\infty$.
2. The root locus of a real system is always symmetrical about the real axis.
3. The system will become unstable when root locus enters the right-half plane.
4. For a realizable transfer function $G(s)$ there are as many root-loci branches as the number of poles of $G(s)$. If the numerator and denominator of $G(s)$ are respectively of degree m and n, there are n-m zeros of $G(s)$ located at infinity, hence n-m branches terminating at infinity.
5. The asymptotic center is the intersection point of all linear asymptotes. For a real system this point is on the real axis

$$\sigma_\infty=\frac{\sum_{r=1}^{n}p_r-\sum_{r=1}^{m}z_r}{n-m} \tag{10}$$

p_r being a pole and z_r a zero of $G(s)$.

6. At a junction point the tangents to the locus are equally spaced over 2π radians.

Property 1 is evident; properties 2 and 3 are rather obvious since they refer to real systems. Property 4 is derived from property 1 by observing that a rational function has an equal number of poles and zeros. Property 5 is derived by observing the behavior of equation 11 when s goes to infinity

$$k\left(s^m+\sum_1^m z_r s^{m-1}+\ldots+\prod_1^m z_r\right)-\left(s^n+\sum_1^n p_r s^{n-1}+\ldots\prod_1^n p_r\right)=0 \tag{11}$$

Example: consider

$$G(s)=\frac{s^3}{(s+1)^2(s+2)}$$

$$1+kG(s)=0 \qquad 0<k<\infty$$

$$\text{or } G(s)=-\frac{1}{k}$$

Following rule 4, the locus has three branches, none terminating at infinity. Based on rule 1, the three branches of locus start from −1, −1, and −2. They terminate at 0, 0, 0. A part of the real axis is included in the locus. Rule 6 gives the 90-degree angle at the junction point −3/2 and 120-degree angles at zero.

To find the breakaway point at −3/2, give an incremental change to an exploratory point in the s-plane about the real axis; see Fig. 4.

$$\text{arg of } G(s)=\text{arg}-\frac{1}{k}=-180°$$

$$\alpha+2\beta-3\gamma=-190°$$

$$\frac{\Delta h}{x}-2\frac{\Delta h}{1-x}+3\frac{\Delta h}{2-x}=0$$

$$x=\frac{1}{2}$$

For $0<k\leq 9$ the system is stable and for $k>9$ two roots enter the right-half plane and the system becomes unstable.

This gives a brief idea of a practical method of finding root locus. Other interesting rules have been devised by servo engineers, which are very useful on account of their engineering applications; see for example the work of Evans.[7–9] The object of the next section is to provide a deeper mathematical insight to these methods. It should be noted that the aim is simply to provide the research scientist in the field of control with the mathematical foundation of the problem and to relate this problem to other problems of mathematical physics. It is not the purpose of this paper to provide immediately applicable methods for the application engineer.

General Properties of Root-Locus Curves

Polynomial Root Locus

Consider an algebraic polynomial of degree p, $P(s)$ in the complex variable s; the coefficients of $P(s)$ may be real or

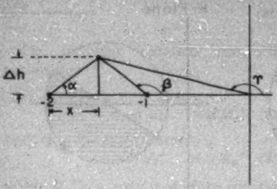


Fig. 4 (left). Determination of junction point

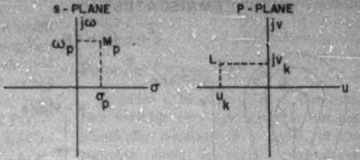


Fig. 5 (right). Conformal correspondence between s and P plane

complex numbers. To a point M_p of the s-plane with the co-ordinate $s_p=\sigma_p+j\omega_p$, there corresponds a point L in the P-plane see Fig. 5. Conversely, to each point L of the P-plane with co-ordinate $k=u_k+jv_k$, there shall correspond a number of points M_p in the s-plane which represent the roots of the polynomial

$$P(s)=k \tag{12}$$

The relation, equation 12, defines a conformal transformation between points of s-plane and p-plane. The root-locus curve is the locus of all root points (M_p) when L describes a given contour of the P-plane.

If roots points M_1, M_2, . . . M_p correspond to a point L with co-ordinate k, when L is moved to location L' with co-ordinate k', the root points take a new position M_1', M_2', . . . M_p'. The following relations are self-evident.

$$(s-s_1)(s-s_2)\ldots(s-s_p)=P(s)-k \tag{13}$$

$$(s-s_1')(s-s_2')\ldots(s-s_p')=P(s)-k' \tag{14}$$

s_k being the co-ordinate of a root point of equation 13 and s_k', the co-ordinate of a root point of equation 14. One concludes that if a variable point of the s-plane assumes the value of a root point, say s_1', then

$$(s_1'-s_1)(s_1'-s_2)\ldots(s_1'-s_p)=P(s_1')-k=k'-k \tag{15}$$

This relation can be very valuable in connection with root-locus problems. Once a set of root points is known, this relation reveals the approximate location for a root s_1' when k is changed slightly. For this purpose, one has merely to observe the approximate magnitude and argument of M_1M_1'. This is shown in the following general theorem.

Theorem 1

If $|M_k|$ represents the root points of a polynomial, $P(s)=k$ for a given value of k. When the value of k is changed to k', the points of the corresponding root locus satisfy

$$\prod_{k=1}^{k=p}\overline{MM_k}=|k'-k|=\overline{LL'} \tag{16}$$

$$\sum_{k=1}^{k=p} MM_k\sigma=\arg(k'-k)=L'Lu \pm \text{multiple of } 2\pi \tag{17}$$

$\overline{MM_k}$ and $MM_k\sigma$ stand respectively for the magnitude and argument (arg) of MM_k, σ being the point at infinity of the positive real axis.

Now that a general relation between the two sets of points of the two planes is established, a particular path must be assigned in the k-plane and the corresponding root locus found in the s-plane. Logically, and also from an engineering viewpoint, interest is mostly in two kinds of contours: infinitesimal path and linear path.

When such studies are conducted, the next logical step is to consider transformation of the infinite contour, as dual of an infinitesimal contour, and a circular path for point k as the orthogonal contour to the linear path. In brief, a mathematical treatment of the subject should use the following root locus corresponding to:

1. An infinitesimal k-contour.
2. An infinitely large k-contour.
3. A circular k-path.
4. A linear k-path.

ROOT LOCUS CORRESPONDING TO INFINITESIMAL K-CONTOUR

In many engineering problems, the point L might move along an infinitesimal contour, say, small changes in the gain of amplifiers or other devices. By the use of theorem 1, it is easy to show that the root-locus curve must form a closed loop around the corresponding root. For a deeper insight into the problem, one may assume that a root point s will move to a location $s+z$ for small change of k.

$$P(s+z)=k+\Delta k \tag{18}$$

Expansion of $P(s+z)$ gives

$$P'(s)z+P''(s)\frac{z^2}{1.2}+\ldots P^{(p)}(s)\frac{z^p}{p!}=\Delta k \tag{19}$$

If $P'(s)\neq 0$, then z also must be infinitesimal. When a point L on the P-plane is close to a location $P'(s)=0$, then

$$z^2\cong\frac{2\Delta k}{P''(s)} \tag{20}$$

Again, the root locus is infinitesimal, but certain topological changes are in sight. This analysis could be further elaborated by theorem 2.

Theorem 2

1. When the point L' moves around a contour which does not include a point corresponding to $dP^{(k)}(s)/ds^k=0$, $k=1, 2, \ldots k$, then the root loci consist of p separate closed curves around root points s_k.
2. When the contour followed by point L' is expanded to include a point L_0 so that

$$\left.\frac{dP^{(k)}(s)}{ds^{(k)}}\right|_{\text{for } s=s_0}=0,\ P(s_0)=k_0 \tag{21}$$

then $k+1$ root loci will be unified in one closed curve.

Without further elaboration on this theorem at present, the importance should be pointed out of focusing general attention on the use of the derivative in allied problems. The author has found it important to consider the notion of derivative in the system analysis and synthesis. Some other instances where interesting results are obtained from the use of the derivative are given by Reza.[13]

ROOT LOCUS CORRESPONDING TO AN INFINITELY LARGE K-CONTOUR

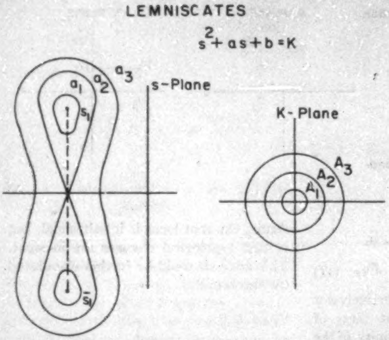
Equation 16 shows that when $\overline{LL'}$ becomes exceedingly large, $\overline{MM_k}$ must also become large. All roots of equation 14 become very large. If units masses are assigned to the roots of equation 12, the center of the gravity G of this system is such that permits the following relations for a point L' far away.

$$\overline{MG}^p=\overline{LL'};\ \text{for } \overline{LL'} \text{ large} \tag{22}$$

$$p\times MG\sigma=LL'u \pm \text{multiples of } 2\pi \tag{23}$$

Equations 22 and 23 are immediately derived from equations 16 and 17 by geometrical consideration. An analytical treatment is also easy to achieve.[8]

It is to be noted that the root points corresponding to $k=\infty$ are the vertices of a p-sided regular polygon; thus, the root loci curves have p real asymptotes symmetrically located about the center of gravity of the roots of equation 12.



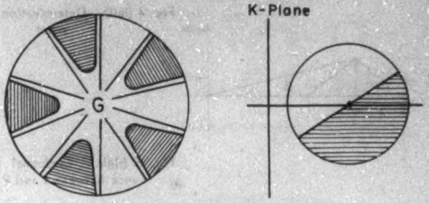


Fig. 6 (left). Example of lemniscates

Fig. 7 (above). A starshaped root-locus

Root Locus Corresponding to Circular k-Path: Cassinians and Lemniscates

In analytical geometry a Cassinian oval is described as the locus of a point which moves in a plane so that the product of its distances from two fixed points of the same plane remains constant. For a moment, assume equation 12 to be of second degree and with real coefficients

$$|(s-s_1)(s-\bar{s}_1)|=|k| \tag{24}$$

From the previous discussion, it becomes evident that for small values of $|k|$, the root loci are two small curves around s_1 and $\bar{s}_1$, and for large values of $|k|$, the root loci become unified in one large contour. The transition from two to one contour starts at a value k_0 which corresponds to s_0, the root of the derivative of equation 24

$$\left.|(s-s_1)(s-\bar{s}_1)|\right|_{s_0=s} = \left|\frac{(s_1-\bar{s}_1)^2}{4}\right| \tag{25}$$

Fig. 6 reveals these considerations (For more detailed discussion as well as the generalization, references 13 and 14 may be consulted.) To turn back to the general equation, equation 12 it is found that the root locus for $|k|$=constant is a generalized Cassinian. These curves are called lemniscates.

$$|P(s)|=|s-s_1|\,|s-s_2|\ldots|s-s_p|=\text{constant}=|k| \tag{26}$$

Some of the properties of these curves have already been discussed; to add a few remarks it may be noted that the branches of one locus or two root loci usually do not intersect. Indeed, for a common point on the two curves, there should be

$$P(s_0)=|k_1| \tag{27}$$
$$P(s_0)=|k_2|$$

except for a multiple point of $P(s)$. For such a point, naturally, two or more branches intersect. It can be shown that for small values of $|k|$, root loci consist of p mutually disjoint branches. Each branch consisting of a Jordan curve around a root point corresponding to $k=0$ (see reference 6, chapter 11). For further discussion, reference 15 may be of interest.

Root Locus Corresponding to a Linear k-Path: Stelloide or Star-Shaped Root Locus

In many problems, it is of interest to find the root locus corresponding to a straight line of the P-plane. Such curves have been called by Lucas "stelloides" (see reference 16, page 17), or star-shaped. The equation of such a curve is given by

$$\arg|P(s)|=\arg k=\text{constant} \tag{28}$$

A star-shaped curve has p hyperbolic branches which are asymptotic to the same previously mentioned asymptote lines. As point L of the P-plane approaches infinity, the root points approach the vertices of a p-sided polygon. When point L moves on the same line but in the opposite direction, the root points thus approach infinity but in directions which form$\sqcap/p$ degrees with the previous asymptotes. Fig. 7 shows such a configuration for a general fifth-degree polynomial. G is the center of gravity of unit masses located at the root points of $P(s)=0$. When the straight line of the P-plane passes through a point L_0, which corresponds to a point s_0 where

$$P^{(n)}(s_0)=0, \quad P(s_0)=k_0 \tag{29}$$

then $n+1$ branches of the root loci will be unified in one branch. This incident is similar to one already discussed for Cassinian curves.

In most engineering problems, a polynomial $P(s)$ with real coefficients must be considered. In such problem k is also generally real and the root locus will correspond to $\arg|P(s)|=0$.

For cases where $P'(s)$ has no real roots, then there are p-1 values of k such that every two are complex conjugates. P-plane will have $(p$-$1)/2$ such points in the upper half-plane and $(p-1)/2$ in the lower. The star-shaped root locus will look like a generalization of Fig. 6. Each half of the P-plane is transformed in $(p+1)/2$ contours, and these $p+1$ contours cover the whole s-plane. In Fig. 6, a large circle was drawn in the p-plane, which is transformed into a large lemniscate (looking like a circle) in the s-plane. It is evident that the lemniscate and star-shaped family of curves considered here form an orthogonal family.

Root Locus for Rational Fractions

In many engineering problems, one is dealing with a system function which appears as a rational fraction $P(s)/Q(s)$

$$\frac{P(s)}{Q(s)}=k \tag{30}$$

As in the third section, consider equation 30 defining as a conformal relationship between points of two planes. Since the trend of the discussion has already been outlined in that section similar detailed discussion may be omitted. The root locus corresponding to a circle of the k-plane will now be studied.

Cyclides

What is the geometric nature of the locus of the roots of equation 30 when point L moves on a circle? To answer this question, a circle of the P-plane may be defined as the locus of points for which the ratio of the distances to two given points L_1 and L_2 is constant

$$\frac{\overline{LL_1}}{\overline{LL_2}}=\text{constant} \tag{31}$$

The following equations are self-explanatory

$$P(s)-k_1Q(s)=MM_1'\times MM_2'\times MM_3'\ldots\times MM_p' \tag{32}$$

$$P(s)-k_2Q(s)=MM_1''\times MM_2''\times MM_3''\ldots\times MM_p'' \tag{33}$$

where M_p' is a point root of the equation 30 for $k=k_1$ and M_p'' a point root of equation 30 for $k=k_2$.

The ratio of these two equations gives

$$\frac{P(s)-k_1Q(s)}{P(s)-k_2Q(s)}=\frac{\dfrac{P(s)}{Q(s)}-k_1}{\dfrac{P(s)}{Q(s)}-k_2} \tag{34}$$

If a point L of the P-plane moves on a circle, equation 31, then $|(k-k_1)/(k-k_2)|$ remains constant, or

$$\frac{MM_1'\times MM_2'\ldots\times MM_p'}{MM_1''\times MM_2''\ldots\times MM_p''}=\left|\frac{k-k_1}{k-k_2}\right|=\text{constant} \tag{35}$$

This is the well-known geometrical property of "cyclides" which was extensively discussed in the 19th century by Darboux, Lucas, Maxwell, and others.[16,17,18]

We may choose another property of the P-plane curve and translate it into property of cyclides. In fact, a circle of the P-plane could be also considered as the locus of points L so that the angle L_1LL_2 remains unchanged.

$$L_1LL_2=\arg(k-k_1)-\arg(k-k_2)=\text{constant} \tag{36}$$

or

$$M_1'MM_1''+M_2'MM_2''+\ldots M_p'MM_p''=\text{constant} \tag{37}$$

Equations 35 and 37 are known to those in the field of servo control by magnitude and angle criteria. The engineering applicarion of these two relations is based on the fact that if a point root is considered, say, close to M_1', then for small variation of k, all the magnitudes encountered in equation 35 remain practically unchanged with the exception of one. A similar statement could be made for the arguments involved in relation 37.

Root Locus Corresponding to Small Variation of k

By an expansion similar to equation 19 it may be shown that if $|P(s)/Q(s)|'\neq 0$ about the points of the root locus corresponding to $k+\Delta k$, then all roots must move infinitesimally. As a result theorem 3 may be stated.

Theorem 3

When a point L moves about a small contour in the P-plane containing no point L_0 corresponding to a critical point of $P(s)/Q(s)$, the root-locus curves consist of a p small, closed curves, each containing only one of the roots of $P(s)-k_1Q(s)=0$. (In the P-plane circle k_1 is included.)

Now, consider a case for which $|P(s)/Q(s)|s'=s_0=0$ and let $P(s_0)/Q(s_0)=k_0$; when point L approaches L_0, the point representing the value k_0, then equation 30 will have two close roots. Indeed, two different branches of a root locus must touch each other at that point and the locus will have a junction point. Similar statements are valid for multiple points of higher orders. Otherwise, in general as two points L_A and L_B move on two separate paths on the P-plane, the corresponding root-loci M_a and M_b do not cross each other.

Root Locus of a Real Rational Fraction for Real k

In servo problems, one is generally dealing with a rational fraction with real coefficients. Furthermore, generally point L moves along the real axis. For $k=0$, root points of equation 30 belong to $P(s)=0$ and for $k=\infty$, to the polynomial $Q(s)=0$. If it is assumed that none of the critical points of $P(s)/Q(s)$ (in recent mathematical literature roots of the derivatives of a function are called critical points) lie on the axis of the reals in P-plane, then the root locus has exactly p branches starting from roots of $P(s)=0$ and ending to roots of $Q(s)=0$, as we move from zero to infinity on the real axis of the P-plane.

Some of the properties of cyclides have already been established which obviously will be valid for this case also. Recently two elaborate doctoral theses related to this particular problem (root locus for real rational fraction when k is real) have been undertaken by Yeh[11] and Chu.[19] Reference 11 has particularly discussed this problem in detail and worked out loci corresponding to a simpler type of rational fractions.

Besides references 11 and 20, there are a number of articles concerned with the behavior of the root locus for real rational fractions. The existence of this literature relieves the writer of further emphasis on this particular problem. Nevertheless, it is noted that the latter literature in most cases is confined to the magnitude and angle criterion. Perhaps a more general approach based on conformal mapping properties along with a classical knowledge of potential curves would be more rewarding for those interested in research in the field of stability and system design.

Potential Curves

Consider a function $Z(s)$ of the complex variable $s=x+jy$, analytic but not identically constant in a region of the complex plane. In such a region, the following relations are self-evident

$$Z(x+jy)=U(x,y)+jV(x,y) \tag{38}$$

$$R^2=U^2+V^2 \tag{39}$$

$$\theta=\tan^{-1}\frac{V}{U} \tag{40}$$

The four functions defined in the foregoing satisfy the familiar Laplace differential equation

$$\Delta U=0 \tag{41}$$

$$\Delta V=0 \tag{42}$$

$$\Delta|\log R|=0 \tag{43}$$

$$\Delta\theta=0 \tag{44}$$

The curves $U(x,y)=\text{constant}$, $V(x,y)=\text{constant}$, form an orthogonal family in the rectangular system of co-ordinates. The same is true for the family $R=\text{constant}$, $\theta=\text{constant}$ in the polar system.

Now, to turn for a moment to the root-locus equation

$$U(x,y)+jV(x,y)=k \tag{45}$$

when, for instance, the point L moves on the real axis

$$V(x,y)=0 \tag{46}$$

and when point L moves on a circle with its center at origin

$$R(x,y)=\text{constant} \tag{47}$$

curves of equation 47 are obviously among the curves satisfying the Laplace equation. This remark suggests that for a deeper knowledge of the root-locus curves or the study of stability in engineering problems, one may profitably investigate the general properties of the curves satisfying the Laplace equation. It is the purpose of this section to give only an introduction to where and how one should start this investigation. The actual study of potential curves will not be undertaken here.

The study of the potential curves goes as far back as the first half of the 19th century, during the third quarter of which the study of level curves or isotherms became a part of mathematical physics. In addition to the previously mentioned references, the work of Klein, LaGoupilliere,[20] Guébhard,[21] and Holzmüller[22] should not be overlooked.

It should be noted that a serious study of the subject was not really undertaken before the turn of the century. It is mostly due to the work of Kasner that some further knowledge of potential curves has become available.[23]

A few typical properties of the potential curves which seem to have immediate bearing on related engineering problems will now be outlined.

PROPERTIES OF ROOT-LOCUS CURVES, CONSIDERED AS POTENTIAL CURVES

If $\phi(x\ y)=0$ is the equation of root locus in the rectangular system, then from the fundamental relation

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}=0 \qquad (48)$$

the following properties might be concluded.

Theorem 4

If ϕ has a finite node, the branches cut at right angles.

Theorem 5

The p asymptotes of a potential curve of the pth order are real, concurrent, and disposed symmetrically about their common point, the angle between consecutive asymptotes being $2\pi/p$. (It may be of interest to note that this theorem already known to Briot, Bouquet,[24] and their students around 1875, has been recently independently rediscovered for its application to servo problems.[11])

Theorem 6

The point of concurrence of the asymptotes of ϕ is the center of gravity of unit masses placed at all the real centers of ϕ.

(For the definition of real center of a curve and the proof see reference 25.) These theorems are quite useful in the study of root-locus curves of some complexity.

It is hoped that the general methods outlined here may be of some use to those interested in basic research in the field of control. The unification of the two families of curves (potential and root-locus) will help in visualizing potential analogue fields, which are rather more familiar to engineers and physicists. In this respect the use of the potential analogue methods may be well worth further investigation. Guébhard in 1882[22] printed pictures of potential curves related to a number of fields of charges. If a short catalogue of similar nature may simplify the solution of many servo problems of wide interest, perhaps it is worth being considered.

References

1. STABILITY OF FLOW IN A ROCKET MOTOR, **D. F. Gunder, D. R. Friant.** *Journal of Applied Mechanics*, American Society of Mechanical Engineers, New York, N. Y., vol. 17, no. 3, 1950, pp. 327–33.

2. STABILITY OF FLOW IN A BIPROPELLANT ROCKET MOTOR, **S. Sherman.** *Ibid.*, vol. 19, 1952, p. 125.

3. ANALYSIS INVOLVING DIGITAL COMPUTERS, **W. K. Linvill, J. M. Salzer.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, July 1953, pp. 901–06.

4. FREQUENCY ANALYSIS OF DIGITAL COMPUTERS OPERATING IN REAL TIME, **J. M. Salzer.** *Ibid.*, vol. 42, Feb. 1954, pp. 457–66.

5. THE GEOMETRY OF THE ZEROS OF A POLYNOMIAL IN A COMPLEX VARIABLE, **M. Marden.** *Transactions*, American Mathematical Society, New York, N. Y., vol. 3, 1949.

6. THE LOCATION OF CRITICAL POINTS OF ANALYTIC AND HARMONIC FUNCTIONS, **J. L. Walsh.** *Colloquium Publication, Ibid.*, vol. 34, 1950.

7. THE USE OF ZEROS AND POLES FOR FREQUENCY RESPONSE OR TRANSIENT RESPONSE, **W. R. Evans.** *ASME Paper no. 53-A-24*, American Society of Mechanical Engineers, New York, N. Y., 1953.

8. GRAPHICAL ANALYSIS OF CONTROL SYSTEMS, **W. R. Evans,** *AIEE Transactions*, vol. 67, pt. I, 1948, pp. 547–51.

9. CONTROL SYSTEM SYNTHESIS BY ROOT LOCUS METHOD, **Walter R. Evans.** *Ibid.*, vol. 69, pt. I, 1950, pp. 66–69.

10. AERODYNAMIC STABILITY AND AUTOMATIC CONTROL, **W. Bollay.** *Journal of the Aeronautical Sciences*, Easton, Pa., Sept. 1951, pp. 569–624 (excellent survey; comprehensive bibliography).

11. THE GENERALIZED ROOT LOCUS METHOD, **V. Chang-Ming Yeh.** *Sc.D. Thesis*, Massachusetts Institute of Technology, Cambridge, Mass., 1952.

12. CERTAIN USES OF DERIVATIVE IN NETWORK THEORY, **F. Reza.** *RLE Progress Report, Ibid.*, Apr., July 1953.

13. SUR UNE CLASSE REMARQUABLE DE COURBES ET DE SURFACES ALGÉBRIQUES ET SUR LA THÉORIE DES IMAGINARIES, **G. Darboux.** Librairie Scientifique Hermann, Paris, France, 1896 (includes good bibliography of work done earlier).

14. LEÇONS SUR LES COORDONNÉES CURVILIGNES ET LEUR DIVERSES APPLICATION, **G. Lamé.** Mallet-Bachelier, Paris, France, 1859, pp. 179–236.

15. LEMNISCATES AND EQUIPOTENTIAL CURVES OF GREEN'S FUNCTION, **J. L. Walsh.** *American Mathematical Monthly*, Buffalo, N. Y., vol. 42, 1935, pp. 1–17.

16. GÉOMÉTRIE DES POLYNOMES, **F. Lucas.** *Journal de l'École Polytechnique*, Paris, France, vol. 28, 1879.

17. PROPRIÉTÉS GÉOMÉTRIQUES DES FRACTIONS RATIONELLES, *Ibid., Comptes Rendus de l'Académie des Sciences*, Paris, France, vol. 78, 1874, pp. 140–44, 180–83, 271–74.

18. ON THE CYCLIDE, **J. C. Maxwell.** *The Quarterly Journal of Pure and Applied Mathematics*, London, England, vol. 9, 1868, pp. 111–26 (historical relation to the subject).

19. A GENERALIZED THEORY OF MULTI-LOOP LINEAR AUTOMATIC CONTROL SYSTEMS, **Y. Chu.** *Sc.D. Thesis*, Massachusetts Institute of Technology, Cambridge, Mass., 1953.

20. THÉORIE DES LIGNES ISOTHERMES, **LaGoupillière.** *Journal de l'École Polytechnique de Paris*, Paris, France, no. 38, 1861 (rather elementary).

21. FIGURATION ÉLECTROCHIMIQUE DES LIGNES EQUIPOTENTIELLES SUR DES PORTIONS QUELCONQUES DE PLAN, **A. Guébhard.** *Journal de Physique, Theorique et Appliqué*, Paris, France, vol. 1, 1882 (use of principle of electrolytic tank produces variety of configuration; bibliography of electrical work).

22. EINFÜHRUNG IN DIE THEORIE DER ISOGONALEN VERWANDTSCHAFTEN UND DER CONFORMEN ABBILDUNGEN, **G. Holzmüller.** Leipzig, Germany, 1882 (still useful for conformal mapping).

23. ON THE ALGEBRAIC POTENTIAL CURVES, **E. Kasner.** *Bulletin*, American Mathematical Society, New York, N. Y., vol. 7, 1901, pp. 392–99.

24. HÖHERE KURVEN, **Briot, Bouquet.** Salmon-Fiedler.

25. ALGEBRAIC POTENTIAL CURVES, **J. L. Coolidge.** *Annals of Mathematics*, Princeton, N. J second series, vol. 27, 1926, pp. 177–86.

A Line Voltage Regulator Having Magnetic-Amplifier Control

F. W. ANDERSON
ASSOCIATE MEMBER AIEE

Synopsis: Closely regulated supply voltages decrease maintenance, improve reliability, and simplify the design of electron tube and semiconductor circuits. There has existed a need for an alternating-voltage regulator which is effective for simultaneous variations in line voltage, load, and frequency, and which has a degree of reliability beyond that associated with electron-tube regulating circuits. Such a regulator, having a 3-kilovolt-ampere (kva) 115-volt 60-cycle rating is described. A power circuit consisting of a saturable reactor and a buck-boost transformer are controlled in a feedback loop by means of a cascaded 2-stage magnetic amplifier. Reference current is provided by a thermistor network and a silicon-carbide varistor is employed as a current preamplifier. These elements are insensitive to frequency over a ±5-per-cent range. The field of application for this type of regulator includes loads within the range of 1 to 20 kva where extremely fast regulation is not essential.

Paper 56-227, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 19, 1955; made available for printing December 20, 1955.

F. W. ANDERSON is with the General Electric Company, Schenectady, N. Y.

The mechanical design of the regulator and the design of the magnetic amplifiers and other coils were the responsibility of R. R. Gay and L. W. Stammerjohn respectively and their associates. C. B. Green and his associates designed the thermistor.

THE need for extremely reliable line voltage regulators capable of compensating for variations in input voltage and load, and which are unaffected by variations in frequency, is well recognized. Unfortunately all of the three major types of voltage regulators used for electronic loads have been compromise designs in the sense that they each have at least one undesirable characteristic. Ferroresonant regulators are sensitive to frequency and, if compensated, are suitable over only a limited load range. Electromechanical regulators are slow in response and are only as reliable as the electron tubes and electronic reference standard in the control circuit. Electronic regulators having a saturable reactor buck-boost power circuit have the same control-circuit disadvantages as the latter.

By replacing the control circuit in the latter type of regulator with a 2-stage magnetic amplifier having a semiconductor reference standard, it has been possible to overcome the aforementioned limitations. This paper describes such a regulator coded *TF161/U* and rated at 3 kva, 115 volts, and 60 cycles.

In addition to the requirement that the regulator regulate for severe variations in input voltage, load, and frequency, its weight was to be about half that of a similarly rated frequency-compensated ferroresonant-type regulator, and its reliability to be such as to preclude the use of renewable components such as electron tubes. Additional requirements included speed of response and output wave form as well as the usual military requirements for shipboard operation. Table I compares the characteristics of the *TF161/U* regulator with those of a 3-kva ferroresonant type employing a frequency-compensating network.

The requirement that regulation be

Table I. Comparison of TF161/U Regulator With a Typical Ferroresonant Type

Specifications	TF161/U	Typical Ferroresonant Type (With Frequency Compensation)
Voltage regulation, ±10-per-cent change in input, per cent	±0.5	±0.5
Load regulation, 10-per-cent load to full load, per cent	1	10
Output voltage variation, ±5-per-cent change in frequency, per cent	±1	±3
Efficiency, per cent: Full load	90	85
Efficiency, per cent: 1/2 load	85	70
Speed of response, 10-per-cent change in input voltage, cycles	12	5
Weight, pounds	175	300
Over-all dimensions, inches	$20\times 23\frac{1}{8}\times 9\frac{3}{8}$	$25\times 23\frac{1}{2}\times 9\frac{3}{8}$

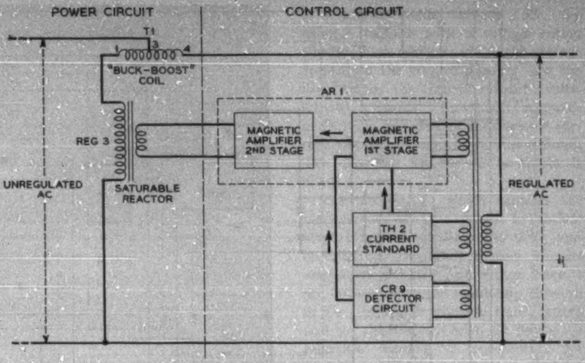
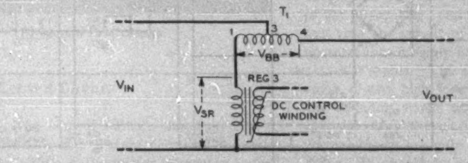
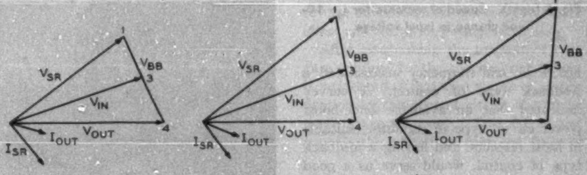

POWER CIRCUIT
CONTROL CIRCUIT
T1
1
3
4
"BUCK-BOOST" COIL
AR 1
REG 3
MAGNETIC AMPLIFIER 2ND STAGE
MAGNETIC AMPLIFIER 1ST STAGE
UNREGULATED AC
SATURABLE REACTOR
REGULATED AC
TH 2 CURRENT STANDARD
CR 9 DETECTOR CIRCUIT

Fig. 1. Block diagram


T1
1
3
4
V_{BB}
REG 3
V_{IN}
V_{SR}
DC CONTROL WINDING
V_{OUT}

(A) TF161/U LINE VOLTAGE REGULATOR POWER CIRCUIT


1
V_{SR}
V_{BB}
3
V_{IN}
V_{OUT}
4
I_{OUT}
I_{SR}

(B) LOW LINE (C) NORMAL LINE (D) HIGH LINE

Fig. 2. Power circuit

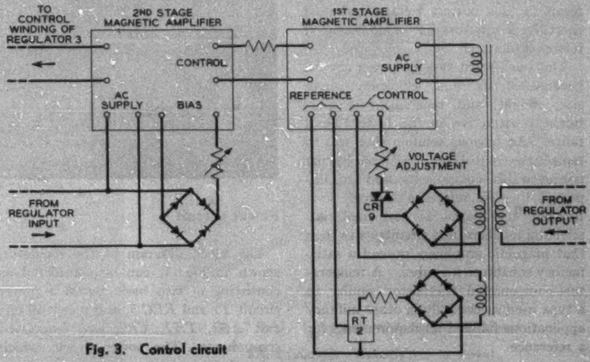

TO CONTROL WINDING OF REGULATOR 3
2ND STAGE MAGNETIC AMPLIFIER
1ST STAGE MAGNETIC AMPLIFIER
CONTROL
AC SUPPLY
BIAS
REFERENCE
CONTROL
AC SUPPLY
VOLTAGE ADJUSTMENT
CR 9
FROM REGULATOR INPUT
FROM REGULATOR OUTPUT
RT 2

Fig. 3. Control circuit

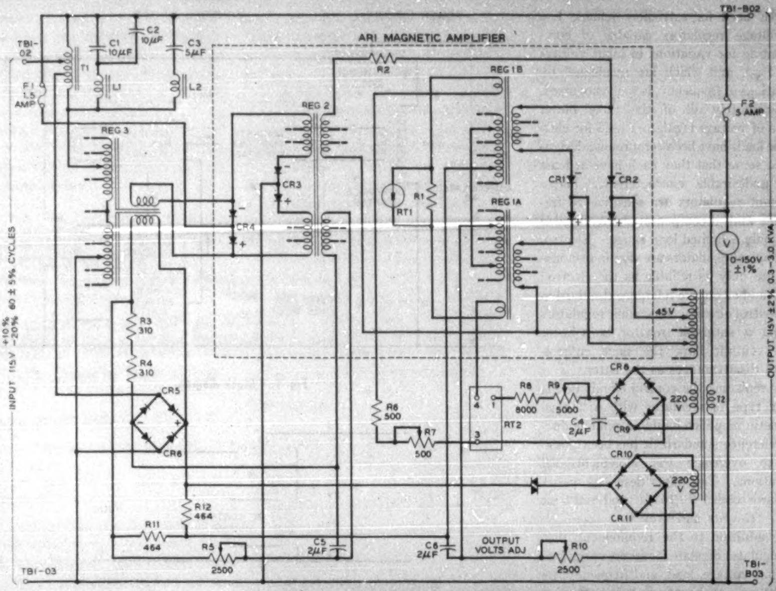
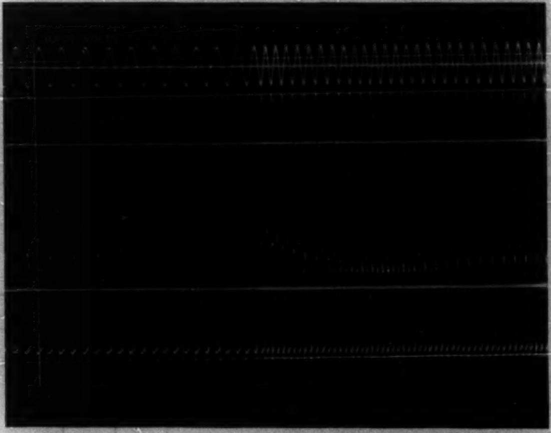

TB1-02
F1
1.5
AMP
T1
C1
10μF
C2
10μF
C3
5μF
L1
L2
AR1 MAGNETIC AMPLIFIER
R2
REG 1B
REG 2
REG 3
CR3
CR4
RT1
R1
CR1
CR2
REG 1A
TB1-B02
F2
5 AMP
V
0-150V
±1%
OUTPUT 115V ±2% 0.3-3.0 KVA
INPUT 115V +10% -20% 60 ± 5% CYCLES
45V
R3
310
R4
310
CR5
CR6
CR8
CR9
220 V
T2
R6
500
R7
500
R8
8000
R9
5000
RT2
4
1
3
C4
2μF
CR10
CR11
220 V
R12
464
R11
464
R5
2500
C5
2μF
C6
2μF
OUTPUT
VOLTS ADJ
R10
2500
TB1-03
TB1-B03

Fig. 4. Over-all schematic

Fig. 5 (right). Speed of response for a −10-volt change in input voltage



age, load, and frequency necessitated a feedback type of control. A survey indicated that an available buck-boost power circuit type of regulator, suitable in most respects, and having a feedback type of control, would serve as a good design point of departure. The significant weaknesses of the available design were its electron-tube control circuit which, because of tube failures, could result in high output voltage and in a short-lived unsaturated diode type of reference detector.

The buck-boost type of power circuit, because of its high efficiency, small size, weight, and reliability, is particularly attractive in the 1- to 10-kva range. An obvious solution then was to replace the control circuit with one having nonrenewable components and an improved long-life reference source. Fortunately the status of core materials selenium rectifiers, and circuitry was such that magnetic amplifiers proved a satisfactory substitute for tubes. A temperature-compensated thermistor similar to a type used successfully in other military applications fulfilled all requirements for a reference.

Power Circuit

The block diagram of the regulator, shown in Fig. 1, can be considered as consisting of two basic parts: a power circuit *T1* and *REG 3*, and a control circuit *AR1*, *TH2*, *CR9*, and associated components. The power circuit, which is not new in concept, consists of a buck-boost transformer and a saturable reactor, as shown in Fig. 2(A). By controlling the impedance of the saturable reactor by varying its d-c control current, the phase and magnitude of the buck-boost transformer current (winding *1–3*) can be adjusted to maintain a constant output

Vol. 7
Pt. 1
1956

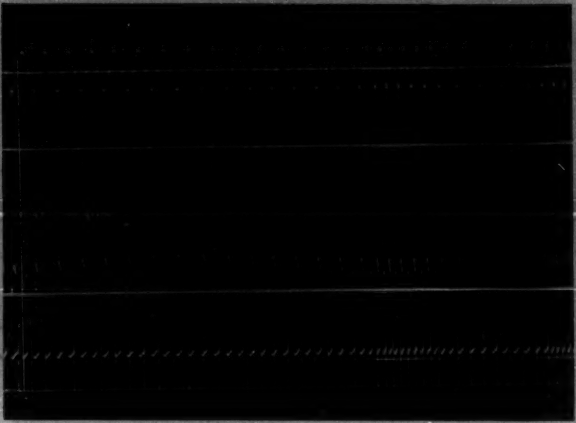


Fig. 6. Speed of response for a −10-ampere load change

voltage. This is accomplished by transformer action through the phasorial adjustment of the input voltage by winding *3–4* on *T1*. This type of power circuit control is phasorially indicated for low-normal-, and high-voltage inputs in Figs. 2(B), (C), and (D). Variations in load magnitude cause almost no change in the voltage relationships indicated because of the current-transformer action of transformer *T1*. As power factor of the load becomes increasingly lagging, the locus of the saturable reactor and buck-boost coil phasors are shifted as they would be for decreasing input voltage.

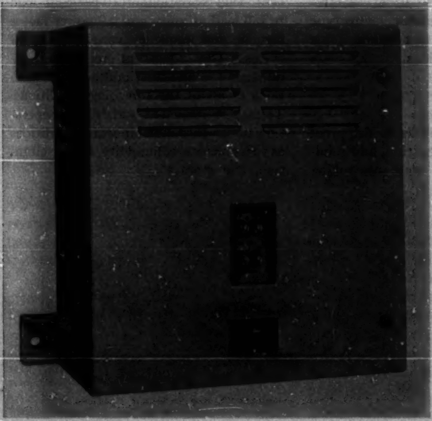
Control Circuit

The control circuit, shown in Fig. 3, consists of a 2-stage cascaded magnetic amplifier, a thermistor current standard *RT2*, a varistor detector circuit current amplifier *CR9*, and associated components. Both magnetic amplifiers are of the bridge self-saturating type with a-c windings in bridge legs common to an a-c input terminal. The two stages, which are mounted as a subassembly, are resistance-coupled. Control signals into the first stage are a rectified sample of the output voltage amplified by means of a silicon-carbide varistor *CR9*. By using a varistor with a voltage rating which is high as compared to the voltage applied to the control winding and the drop across the control resistor (volts-adjust), a current gain equal to approximately the exponent in the varistor current voltage equation $I=KE^N$ is realized. The exponent for the *CR9* varistor is about 4.

Thermistor *RT2* which supplies the the reference and bias current to the first stage magnetic amplifier has three ambient-temperature-compensating circuits associated with it. Two of these circuits are encased with the regulating network in the thermistor epoxide resin casting and the third is cast with the first-stage reference winding to compensate for variations in its resistance.

The output of the control circuit is coupled directly into the control winding of the power stage reactor. Despite the fact that the output of the second stage operates into a highly inductive load, the circuit is stable. This is probably so because of the high forward resistance of the selenium stacks used in the second magnetic-amplifier stage and because of the relatively low gain of the power reactor. It should be said that in both stages the selenium cells were operated at about 1/3 of their voltage rating to limit reverse current which has a degenerative effect and causes long-range instability.

After establishing the amount of loop gain required for the specified regulation, it was decided that two stages of amplification in the control circuit should be used rather than one. Two stages also provide the advantage over one stage of



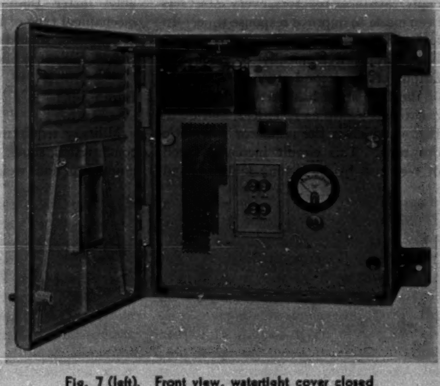
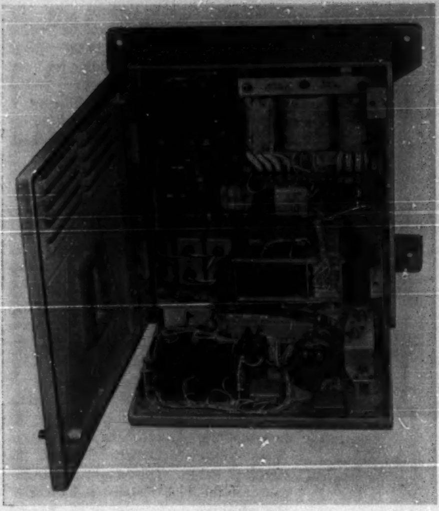


Fig. 7 (left). Front view, watertight cover closed

Fig. 8 (above). Watertight cover open



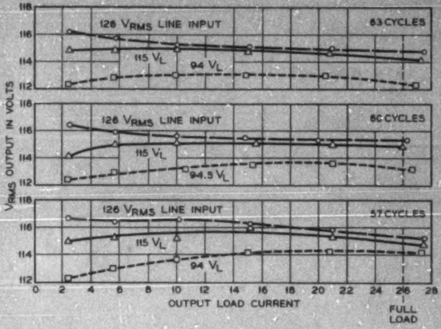


Fig. 9 (left). Watertight cover and hinged control panel open

Fig. 10 (above). Regulating characteristics

faster over-all response. Because only moderate gain per stage was required, magnetic-amplifier core materials, which are commercially more practical than high-nickel-content cores, were employed. The first-stage amplifier cores are grain-oriented silicon-steel tape-wound toroids cast in an epoxide resin casting. The second stage is tape-wound grain-oriented silicon-steel cut cores as are most of the other coils. All coils are insulted with class-*H* insulation.

Because of the inherent relatively slow response of magnetic amplifiers and especially saturable reactors, every effort has been made to improve response time. In addition to the 2-stage amplifier, a feedback circuit dynamically coupling the output of the second stage to the input of the first stage has been added. The network is shown as *R11*, *R12*, and *C6* in Fig. 4, which is the regulator schematic drawning. This circuit increases speed of response by increasing stability and preventing overshoot. To this extent it also serves as an antihunt circuit.

An oscillogram indicating speed of response for a −10-per-cent variation in input voltage is indicated in Fig. 5. The effect of a load variation of −10 amperes on output voltage is shown in Fig. 6. Response of the magnetic amplifiers, as evidenced by their output currents, are also indicated in these figures. These oscillograms establish the fact that for input voltage variations the response of the *TF161/U* regulator is not quite as fast as a ferroresonant type but considerably faster than the usual electromechanical type of regulator. This type of regulator should be fast enough for a wide variety of applications.

Equipment

Military requirements as to dripproof construction, shock, vibration, and minimum size and weight were controlling factors in equipment design. Fig. 7, which shows the cabinet with the cover closed, exhibits the inverted louvers which with holes in the bottom of the cabinet, provide chimney-effect convection cooling. Fig. 8 shows the voltage-adjust potentiometer output voltmeter and perforations in the front panel which assist in selenium stack cooling. In Fig. 9 the magnetic-amplifier subassembly is shown in the bottom right side of the cabinet. The thermistor is shown in the bottom left corner of the open panel. The regulator is suitable for bulkhead mounting and will operate unattended in ambient temperatures ranging from 0 to 50 degrees centigrade.

Performance

Fig. 10 indicates the fact that the regulator meets the over-all requirement of ±2-per-cent regulation for input voltage variations of +10 and −20 per cent, load variation of 10-per-cent load to full load at unity and 0.9 power factor lagging, and frequency variations of ±3 cycles. The 2-per-cent regulation requirement is met for simultaneous variations of all of the foregoing conditions. Tests also indicated satisfactory operation under extreme shipboard conditions as to temperature, humidity, shock, vibration, and storage.

A 72-Channel Radio System for Toll Telephone Service

M. C. HARP
NONMEMBER AIEE

M. H. KEBBY
NONMEMBER AIEE

Synopsis: A multichannel radio system is described which has been developed to meet the specific requirements of toll telephone service for transmitting medium channel loads over medium distances. Problems relating to reliability of service, noise, fading, cross-talk, interconnection and other requirements important to the telephone industry are discussed.

Application

THERE has been a growing demand in the past few years for medium-channel-capacity microwave equipment suitable for use in toll telephone service. Requirements for large-capacity long-haul service are well met by the existing *TD-2* equipment, which is widely used on the nationwide long-lines network. A number of small-capacity systems are available which, although finding considerable use in the industrial field, have only a limited application in the telephone industry. The system described here, designated as type *72B*, is intended to fill the gap between the existing large and small systems. It is tailored to the specific requirements of the telephone toll plant for the many medium-length routes where required channel capacity is measured in dozens rather than hundreds.

The requirements of radio equipment designed to be used for toll transmission in the telephone plant are largely determined by the requirements of the land lines with which the radio equipment must co-ordinate. These are: 1. reliability of the radio equipment should equal that of a wire line along a comparable route, 2. noise and crosstalk should meet standards established for toll circuits, 3. channels transmitted over the radio system must be integrated easily into the existing wire line plant, 4. engineering of a radio installation should be possible with no more effort than is normally required for a comparable wire line system, 5. the radio equipment should accept any of the commonly used wire line carrier systems with a minimum of special interconnecting equipment. These requirements may be summarized by stating that it is desirable to be able to treat the radio circuit as though it were just another wire line facility.

Radio's field of application has grown rapidly in the past few years, partly because of economic considerations which favor radio in many instances, and partly because the industry has come to know and appreciate the advantages of radio as a transmission medium. In many cases radio is providing lower cost connections: natural barriers such as mountains and water obstacles can frequently be spanned much more efficiently by radio circuits. In other situations radio is being used because of its greater reliability: a radio link is subject to less outage time and is easier to maintain than is a wire line through sleet or hurricane regions. Completely new circuit requirements to serve military installations or oil or mining developments are a natural application for radio.

In these situations right-of-way procurement and pole line installation costs are eliminated. In telephone plant, radio systems have even proved in on routes paralleling existing wire line leads where no particular problem existed beyond simply adding a new crossarm on existing poles. The centralized maintenance of a radio system is a very important advantage as compared with a submarine cable or a wire line constructed over rough or inaccessible terrain.

Frequency

For practical purposes, the frequencies available to the telephone common carrier services for multichannel toll transmission within the United States are limited to bands in the 900-, 2,500-, 4,000-, and 6,000-megacycle (mc) regions. Although higher frequencies have also been set aside for common carrier services, their use has been avoided because of the serious problem of precipitation attenuation over 6,000 mc. Large channel groups, which naturally require wide radio channels, at present utilize only the 4,000-mc band. Early use of the 6,000-mc band for large channel groups can be expected.

Type-*72B* equipment operates in the 900-mc band which provides adequate spectrum for medium channel loads. Choice of this frequency band was influenced by a number of factors including maintenance considerations, propagation characteristics, and equipment economies. These factors are of particular importance for the type of service contemplated for this equipment and may not be equally valid in selecting operating frequencies for systems of larger capacity. Equipment operating at 900 mc utilizes circuitry and techniques developed to such an extent for high-frequency and very-high-frequency service that the equipment can be operated and maintained by radio technicians without specialized training in the microwave art.

Fading at 900 mc is generally less severe than at the higher frequencies. As a rough approximation, it is believed that the necessary fade margin varies logarithmically with frequency. Generally, tests over any one given path at various frequencies show that the depth of fades seems to vary in decibels (db) at a rate of somewhat less than 10 log F_h/F_1; however, because the added susceptibility of the higher frequencies to fading caused by rain attenuation and effects resulting from the narrower beam width commonly used which result in increased sensitivity to duct propagation and fog bank refraction, it seems reasonable for practical purposes to figure that the necessary fade margin may vary in db as about 10 log F_h/F_1.

If proposed rule makings of the Federal Communications Commission are finally adopted, the band of frequencies from 890 to 936 mc will be assigned for use by the telephone common carrier service only. This band of frequencies can be divided readily into a total of 18 channels for *72B* equipment. This permits operation of three parallel duplex radio circuits in each of two directions from a central hub location, even where overshoot of signal to the station beyond the next repeater location is encountered. A maximum of 432 telephone channels transmitted over six duplex radio circuits can be terminated in one location. In Fig. 1 is shown a frequency allocation for three parallel duplex circuits which is used for a system extending from Eugene, Oreg., over mountainous terrain to several coastal locations. Even more channels can be terminated at a single location where the overshoot problem can be avoided by judicious location of repeater sites.

Paper 56-16, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted July 19, 1954; made available for printing October 31, 1955.

M. C. HARP and M. H. KEBBY are with the Lenkurt Electric Company, Inc., San Carlos, Calif.

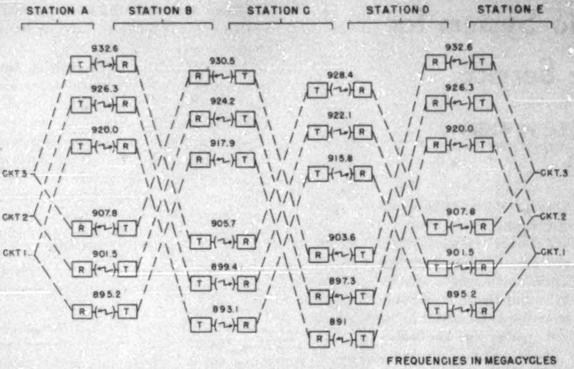

STATION A
STATION B
STATION C
STATION D
STATION E
932.6
926.3
920.0
907.8
901.5
895.2
930.5
924.2
917.9
905.7
899.4
893.1
928.4
922.1
915.8
903.6
897.3
891
932.6
926.3
920.0
907.8
901.5
895.2
T
R
CKT 3
CKT 2
CKT 1
CKT. 3
CKT. 2
CKT. 1
FREQUENCIES IN MEGACYCLES

Fig. 1. Frequency allocation for three duplex type-72B radio systems operating along same path

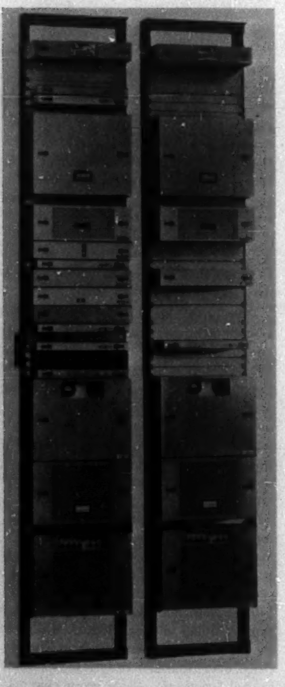


Fig. 2. Two typical duplex equipment assemblies. Radio transmitter, receiver, and associated equipment mounted on each rack. Assembly on left includes auxiliary equipment for supervision and control

Radio Equipment

Two rack bays of type-*72B* equipment are illustrated in Fig. 2. Each rack is equipped with transmitter, receiver, and a-c power supplies, and one rack includes auxiliary equipment for supervision and control. The radio equipment is rack-mounted and equipped with dust covers in the same manner as for widely used types of telephone carrier systems to conform with standard toll office layouts. Every effort has been made in equipment arrangements to provide systems which can be operated and maintained by personnel with a minimum of training and experience in radio techniques.

Block diagrams of the radio transmitter and receiver are shown in Fig. 3. The modulator, converter and power output panels for the transmitter, and the demodulator and converter panels for the receiver have been developed for the type of wide-band low-distortion operation required of multichannel systems used for toll telephone service.

The Serrasoid modulator used in the transmitter allows a high signal-to-noise ratio with extremely low distortion characteristics. It provides a crystal-controlled frequency source and has no circuitry requiring critical adjustment. The modulator provides a phase-modulated signal which is converted to frequency modulation through a 6-db-per-octave corrector (slope) network. Phase modulation is accomplished by clipping a saw-tooth wave form at various points along its slope, the point of clipping being dependent upon the instantaneous amplitude of the modulating signal. Since the slope of the saw-tooth wave form may be made exceptionally linear, amplitude distortion is virtually negligible.

Low-frequency signals from the Serrasoid modulator are translated to a higher frequency, amplified, and multiplied to the final frequency in the 900-mc band. Rated peak frequency deviation is 500 kc. Ordinary negative grid-type tubes are used throughout the equipment, and a *2C39A* triode (lighthouse tube) is used in the power output stage. This tube, rated for up to 100-watt dissipation, is operated at about a 20-watt input to yield an output of from 5 to 8 watts. Users are consistently obtaining useful tube life in excess of 5,000 hours when

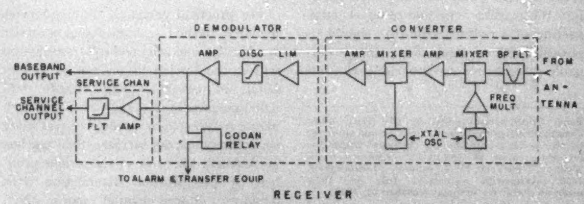

MODULATOR PANEL
CONVERTER PANEL
POWER OUTPUT PANEL
AMP
MOD
TPLR
AMP
MIXER
AMP
TPLR
TPLR
DBLR
BP FLT
BASEBAND INPUT
TO AN TENNA
SERVICE CHAN
FLT PNL
SERV CH PNL
SERVICE CHANNEL INPUT
XTAL OSC
XTAL OSC
TRANSMITTER
DEMODULATOR
CONVERTER
AMP
DISC
LIM
AMP
MIXER
AMP
MIXER
BP FLT
FROM AN-TENNA
BASEBAND OUTPUT
SERVICE CHAN
SERVICE CHANNEL OUTPUT
FLT
AMP
FREQ MULT
CODAN RELAY
XTAL OSC
TO ALARM & TRANSFER EQUIP
RECEIVER

Fig. 3. Block diagram, type-72B transmitter and receiver

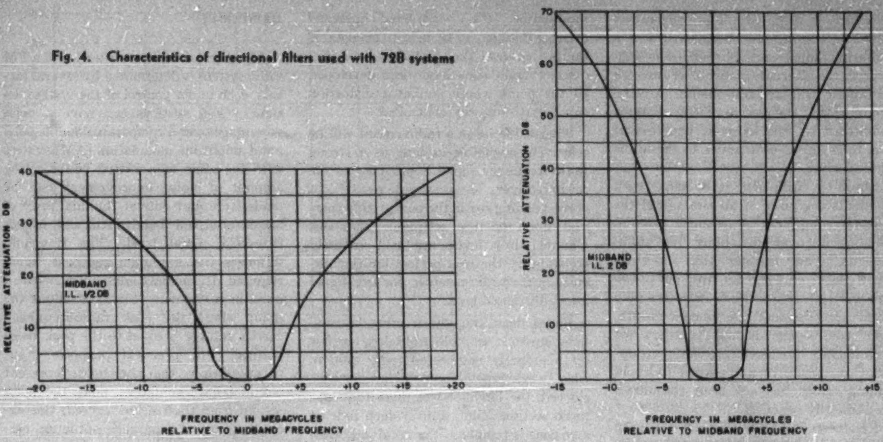


Fig. 4. Characteristics of directional filters used with 72B systems

operating in this conservative manner.

A straightforward double superheterodyne receiver circuit is used with a silicon crystal as the first mixer. Receiver band width is approximately 2 mc to accommodate more than 99.9 per cent (%) of the total side-band power for the rated peak deviation of 500 kc, and about 99% of the side-band power for a deviation of 800 kc. This is equivalent to less than 30-kc bandwidth per channel for a full 72-channel system, a comparatively good usage factor for a 900-mc microwave system. The receiver includes a Codan or squelch circuit which mutes the receiver noise output in the absence of a received signal in much the same manner as is common in the mobile radio services.

Ten-foot mesh parabolic reflectors are most commonly used with *72B* systems although smaller sizes can be employed under favorable operating conditions where less antenna gain is required. A directional filter arrangement included with equipment assemblies permits the use of a single antenna for both transmitting and receiving. With these filters the cross coupling from transmitter to receiver is controlled to lower limits than would be the case if two separate antennas were used without filters. Various directional filter arrangements are used to permit the use of one antenna for up to three receivers, up to three transmitters, or two receivers and two transmitters. The filters consist of 1/4-wave coaxial cavities connected in cascade to provide the necessary suppression to adjacent radio channels. Characteristics of typical transmitting and receiving directional filters are shown in Fig. 4.

Repeaters

Two general types of repeaters have been commonly used in microwave systems.

One type, known as a back-to-back repeater, simply connects the output of a receiver to the input of a transmitter.

The second type, known as an intermediate-frequency or heterodyne repeater, converts the received signals to an intermediate frequency band for amplification. The amplified signals are then converted to a new transmitting frequency.

Although a heterodyne repeater does not introduce the nonlinear distortion associated with modulation and demodulation, it has the serious disadvantage for telephone applications of providing no ready access for dropping or inserting channels at repeater points. Since the need for channel dropping and insertion is quite frequent in short to medium haul toll circuits, the *72B* system has been designed to use back-to-back repeater arrangements. Through conservatively engineered circuits and such techniques as the Serrasoid modulator, the amount of distortion associated with modulation and demodulation has been reduced to a point where about ten repeater sections may be connected in tandem before distortion noise contributions of the radio equipment become the limiting factor in system engineering. Since many more than ten sections are readily practical when compandors are used with the carrier equipment, *72B* systems can be extended well beyond 500 miles if desired.

Supervision and Stand-by Arrangements

Associated with type-*72B* radio equipment is a supervisory, control, and alarm system designated type *51B*. This system utilizes coded tone transmission between a control center and a number of remote stations to actuate a wide variety of alarm, supervisory, and control equipment. Because each radio installation presents different control problems, the *51B* system has been designed on a flexible basis to permit a variety of arrangements. Information concerning the activation of a supervised alarm or transfer function at a remote point will be transmitted automatically to the control center. With stand-by and transfer equipment provided, a failure is immediately corrected and the alternate equipment transmits full information as to what the automatic equipment has done. In the event of a total failure of a radio section, major alarm is given, but the information as to which end of the section is dead can be sent to only one end of the system by the supervisory system. Personnel at the control center can verify the status of all remotely supervised functions immediately by observing a lamp field display. Remote control of the distant equipment is provided from the control points by dialing a coded signal. The *51B* system can supervise up to 80 remote functions and can control up to 90 remote activities divided among up to eight remote locations.

A complete duplicate radio system is provided as stand-by equipment for any location where an unusually high degree of reliability is required. The transmitted power is continuously monitored and a radiation failure causes automatic switchover to the alternate transmitter through automatic transfer equipment. A tone originating at the preceding transmitter is monitored by the receiver, and loss of this tone causes switchover to the alternate receiver.

Stand-by equipment is operated with filaments energized at all times, and the transfer equipment is arranged to complete switchover in about 150 milliseconds. This transfer time is sufficiently fast to prevent the drop-out of toll circuit connections. Since transfers may occur as infrequently as once a month, even considering manual transfers for maintenance purposes, providing equipment to transfer rapidly enough to avoid dialing errors has not been considered economically practical.

Reliability in wire line service is usually assured by providing alternate wire pairs and alternate routes. It is therefore not usually necessary to back up wire line installations with full stand-by alternate equipment. Similarly, when a radio system is part of such a network, stand-by provisions would not ordinarily be required. In cases where no alternate route to the radio system is available, absolute continuity of service may be required. With stand-by equipment and automatic transfer, the radio system is highly reliable with circuit continuity limited only by the probability of fading. This fading can be controlled by careful and conservative engineering of the system.

Noise

Noise in radio links is of two distinct types, each originating from a different source, and each must be handled as a separate problem. One type, resulting from the thermal noise at the receiver input, is heard by the listener as a constant hiss. The second type, caused by nonlinear distortion in the radio equipment, is called "babble" and is similar in many respects to unintelligible crosstalk in a wire line system. The characteristic hiss or rushing sound of thermal noise is less disturbing than distortion noise. Since babble results from distortion products falling into other than the disturbing channel, it creates a noise problem which does not exist in a single-channel radio system.

A frequency-modulated system such as the *72B* provides a means of reducing the thermal noise by choosing a deviation ratio which yields some nominal frequency modulation (FM) wide-band improvement. Babble can be held to acceptable limits because the techniques of frequency modulation have been developed to the point where nonlinear distortion can be satisfactorily controlled.

Thermal noise in a radio circuit will be relatively negligible as long as a strong radio frequency signal is provided to the radio receiver. A signal fade results in a corresponding rise in the background thermal noise on the telephone channels. Channel drop levels are held virtually constant by the near-perfect limiting circuits in the radio receiver, for any signal above threshold level.

Fades deep enough to bring thermal noise up to noticeable audibility are few on a properly engineered radio system. Also, when repeatered systems are considered, the likelihood of more than one radio section falling into a deep fade at any time is remote. The total fade time should be expected to amount to only minutes per month, and on this basis the noise for maximum fade is permitted to rise to a high value. In contrast, distortion-caused noise is contributed simultaneously by all repeater sections and is fairly continuous during busy periods. A much lower level of distortion noise must therefore be maintained.

Signal-to-noise ratios as poor as about 30 db are acceptable as the bottom limit for a deep fade. (The signal-to-noise ratio quoted compares a zero-volume-unit talker in a channel to a 3,000-cycle noise bandwidth in the channel.) This corresponds to per-channel noise of about 52 db adjusted (dba) at zero level. Such noise value is acceptable with the limitation that is should not be attained more than about 0.01% of the time, or about 4 minutes per month. This noise limit is set at about the point where dialing circuits may be affected by noise hits, or where supervision may be interrupted long enough to cause a disconnect. Thermal noise during normal propagation periods is negligible for all practical purposes.

Exact predictions of fades to be expected for a given repeater section are difficult to make. However, it has proved practical to approximate the outside limit of the fade and engineer the system to these requirements. For type-*72B* equipment, fades up to about 44 db can be tolerated over a typical span of about 40 miles without exceeding the specified maximum noise. This large margin is achieved with a moderate transmitter power of about 5 watts and with relatively high antenna gains of about 27 db each.

Bandwidth

The thermal noise contribution of a FM radio system is determined by several factors, each under control of the system designer. FM systems can have a noise advantage over a comparable double side-band amplitude modulation (AM) system with the same transmitted power. The amount of noise improvement can be made large, but only at the dual price of larger occupied bandwidth and higher threshold signal level. The threshold signal is the minimum received signal required to achieve detection with a noise improvement; occurs at about the point where the peak radio-frequency carrier voltage is equal to the peak noise voltage. For practical purposes input signals below the threshold represent circuit failure conditons, since the noise output from the receiver exceeds the signal output. To provide optimum performance at the bottom of a fade, an FM radio system should be designed so the threshold is just beyond the point where noise reaches the target of 52 dba at zero level, as discussed. This provides maximum FM noise improvement with the greatest possible fade margin. In *72B* equipment a receiver bandwidth of about two mc coupled with a receiver noise figure of about 14 db provides an effective compromise between noise improvement and threshold signal level.

Emphasis

In a pure FM system, noise amplitude at the output of the receiver decreases from the high end of the baseband toward the low end at a rate of 6 db per octave. To compensate for this inherent characteristic, equipment usually includes an emphasis system to equalize the noise across the baseband. A pre-emphasis network in the transmitter shapes the baseband so that modulation from the low-frequency channels is decreased at about the same rate as the noise decreases. A curve showing the deviation as a function of frequency is shown in Fig. 5. In other words, more of the total frequency deviation is caused by the high-frequency channels than by those of low frequency. A complementary de-emphasis network at the receiver restores the baseband frequencies to a flat spectrum.

Loading

When used with usual single-side-band frequency-division multiplex equipment, the load fed to the radio transmitter consists predominantly of a group of speech

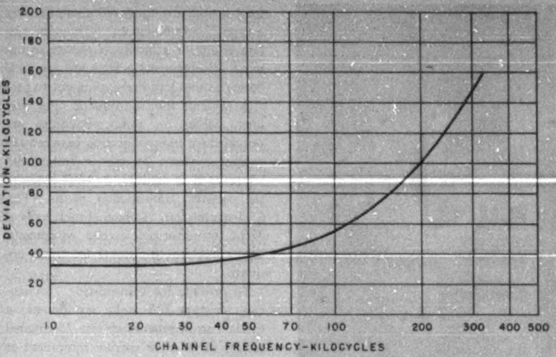


Fig. 5. Deviation of zero dbm tone at zero level as a function of frequency

side-bands. The complex wave form of this modulating signal has a relatively high peak to average ratio, but the highest peaks occur only occasionally during the busiest hours. Detailed study of the statistical properties of a group of single-side-band speech channels has been made by Holbrook and Dixon[1]. In telephone terms, this source indicates that when 72 channels are put through a common transmitting device such as an amplifier, FM transmitter, or receiver, etc., and when the output level is zero db, then the output or load capacity of the device on peak signals should be approximately 18 db above 0.001 watt [18 db above 1 milliwatt (dbm)]. Zero-db level is the reference level at the 2-wire input to toll circuits (toll testboard level). A device at zero-db level output operates with zero-db gain between its output and the toll circuit input. The loading of the system must be set so that these occasional peaks do not cause excessive frequency deviation resulting in objectionable intermodulation noise or babble. Unless excessive deviation is held to short periods, the noise limits will be exceeded. On the other hand, if the system is adjusted for less than the maximum possible deviation, part of the fade margin is sacrificed.

The emphasis treatment used in the *72B* system changes the system loading considerations by modifying the signal peaks of the 72-channel group. Analysis shows that this modification results in a reduction of the peak load factor by about 3 db. Therefore, emphasis permits a net loading of the system of about 3 db higher than would be possible without emphasis. With this information, it has been found that if the highest frequency channel (at 300 kc) causes a deviation of ±150 kc (with a zero dbm signal), the peaks will exceed a deviation of ±800 kc for about 1% of the time during the busy hour. Since each peak lasts for only 1 or 2 milliseconds, distortion resulting from the peaks will not be apparent to the listener, even though the total overload time approaches 36 seconds in 1 hour.

It may be noted that the rated maximum deviation of the *72B* system is ±500 kc, although a deviation of ±800 kc is reached on 1% busy hour peaks. The ±500-kc deviation figure is one which has been used in equipment design including tuned circuit bandwidth requirements in order to achieve acceptable performance most of the time. In contrast to the distortion created in negative feedback amplifier circuits commonly used in carrier equipment, which increases rapidly beyond the overload point, the distortion created in the *72B* system rises gradually under overloaded conditions and does not suddenly increase as in the case of the negative feedback amplifier. For this reason, considerable overdeviation for short-time intervals may be accepted by the *72B* system.

FM Improvement Factor

The per-channel thermal noise improvement of a multichannel FM system in which the noise spectrum has been corrected by emphasis has been shown to be related directly to the square of the per-channel deviation ratio; see the Appendix. This improvement refers to the difference in noise in a channel transmitted over an FM system as compared with a standard reference of a single-channel double-side-band AM system with the same carrier power and with 100% modulation.

In the *72B* system, a zero dbm test tone at zero transmission level placed in the top channel (300 kc) causes a frequency deviation of ±150 kc. This is a deviation ratio of one-half for the system's worst channel. The improvement is therefore $(1/2)^2 = 1/4$, or −6 db as compared with the 100% modulated AM double-side-band system. For purposes of this comparison it is assumed that a zero dbm tone (at zero level) produces 100 per cent modulation of the AM transmitter. Since it has been shown that a test tone of +18 dbm (at zero level) presents loading equivalent to that of 72 channels, it follows that the per-channel level of the 72-channel AM system must be reduced to −18 dbm to present an equivalent load of 0 dbm. This results directly in an 18-db degradation of per-channel signal-to-noise ratio. Since the *72B* system improvement factor (per channel) is −6 db referred to a 100% modulated AM system and the 72-channel AM system must have a per-channel modulation factor 18 db below 100%, it follows that the *72B* effective noise improvement per channel is 12 db above that of a comparable double side-band AM system of the same carrier power.

Distortion

On toll telephone circuits the level of noise caused by nonlinear distortion is expected to be held to very low levels. Such circuits are repeatedly degraded by the tandem connection of many repeater sections to form links and by the ultimate tandem connection of several links to complete a toll circuit. For this reason, one of the targets in the design of a radio system is that a single radio repeater section should contribute only about 1% or 2% of the total noise objective for the end-to-end toll circuit. To meet this objective the channel signal-to-noise ratio in a single radio section should be above 60 db; i.e., the average noise should not exceed 22 dba (*F1A* weighting at zero level) during the busy hour. This requirement can be restated in terms commonly used by radio engineers to state that any product from a zero dbm tone at zero level, or cross product of two such tones, should not exceed about 0.1% distortion for a system such as the *72B*. Experience indicates that radio equipment adjusted to the use of this procedure will yield satisfactory operating results.

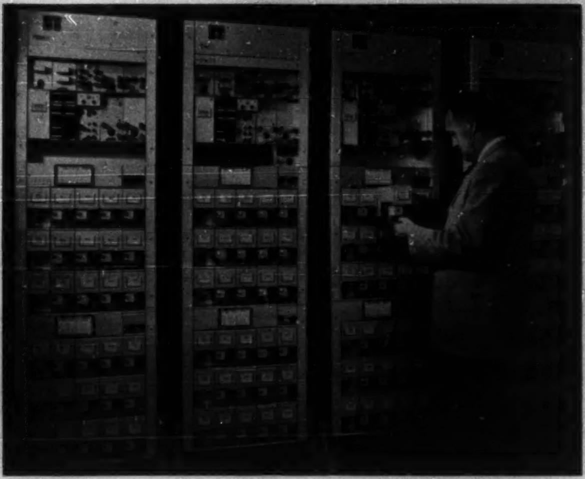


Fig. 6. Two 48-channel terminals, type-45BX carrier telephone equipment

Typical Link

A typical line of sight section for a *72B* system might be as long as 40 miles. At this distance the space loss between antennas should be approximately 127 db. With total antenna gains of about 54 db and about a 4-db attenuation in transmission lines, the net loss between transmitter and receiver would be about 77 db. With a transmitter output of +37 dbm (5 watts), the received signal level would be −40 dbm (2,200 microvolts in 50 ohms). Since the receiver FM improvement threshold is at about −84 dbm, at which point noise in the worst channel is 52 dba, about 44 db remains as the absolute maximum fade depth which can be tolerated. This amounts to a fade margin of over 1-decibel per mile, a value considered adequate for all but the most adverse locations. In an area where fading is expected to be especially severe, the length of the radio span should be reduced to about 25 to 30 miles in order to provide a fade margin of $1^1/_2$ db or more per mile.

Carrier-Channelizing Equipment

Almost any of the commonly used wire line carrier systems can be used for channelizing type-*72B* radio equipment. Although not usually used for groups of fewer than 12 channels, the system has found some application for extending 12-channel cable carrier systems across water barriers.

A recently developed carrier system, the *45BX*, has been specifically adapted for transmitting up to 48 channels over a type-*72B* radio circuit. This new system is one of a family of carrier equipment arrangements which use identical basic channel terminal equipment to provide systems for transmission over wire lines, cables, or radio. Facilities are provided for interconnecting between the various systems at carrier frequencies wherever it is desirable to shift between the various transmission media. Fig. 6 illustrates four 24-channel rack bays of *45BX* miniaturized carrier equipment. Two complete 48-channel terminals are shown.

To provide for transmitting 72 channels on type-*72B* radio equipment, a special arrangement of six 12-channel groups of *45*-class carrier equipment is being developed, utilizing frequencies from 12 to 300 kc.

Installations

About 25,000 channel miles of *72B* equipment will have been installed to provide toll telephone service by the end of 1954. The equipment is being used to meet a variety of problems. In one case a radio link has provided circuits between Seattle, Wash., and Victoria, B. C., to replace previous facilities which included a circuitous land line and a failing submarine cable. Other installations are being made in locations throughout the mountainous terrain of the Pacific coastal area, where wire line failures have been commonplace during severe winter weather. Plans are also being made to replace existing wire line sections in other regions where land line reliability has been a problem in the past. Several installations are under way to provide communication service to new military bases remote from existing wire line facilities. A number of systems are being used in the Hawaiian Islands where submarine cables have not been practical for inter-island circuits.

Perhaps the most significant of the type-*72B* installations are those being made where no particular barriers exist,

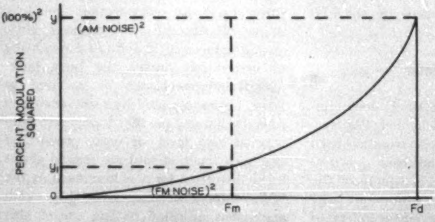


Fig. 7. Wide-band FM systems. F_m = highest modulated frequency. F_d = deviation of highest modulated frequency. D = deviation ratio

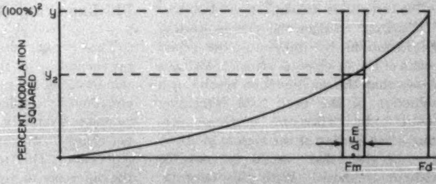


Fig. 8. Channelized FM systems

or where the capacity of existing land lines could be increased with additional crossarms. Several such installations have been or are being made, in some cases because radio systems can be installed more rapidly than a pole line, but usually because economic studies have shown that the additional channels will cost less, will be easier to maintain, and more reliable if radio is used. There can be no doubt that the next few years will witness a marked swing to the use of radio in the normal growth of short and medium haul toll circuits in the telephone industry.

Appendix

The following shows the development of a simple method for obtaining the signal-to-noise ratio and the FM-AM improvement factors for both wide-band and channelized FM systems. Fig. 7 shows the noise spectrum for both FM and AM systems. The AM noise is proportional to the noise contained below the dashed rectangle whose area may be designated

$$(N_{AM})^2 = yF_m \tag{1}$$

The FM noise is proportional to the area under the parabolic curve up to y_1 and may be designated

$$(N_{FM})^2 = \frac{y_1F_m}{3} \tag{2}$$

Also

$$y = kF_d^2 \tag{3}$$

$$y_1 = kF_m^2 \tag{4}$$

$$\frac{y}{y_1} = \frac{kF_d^2}{kF_M^2} \tag{5}$$

By definition

$$D = \frac{F_d}{F_m} \tag{6}$$

Substituting equation 6 in equation 5

$$\frac{y}{y_1} = D^2 \text{ or } y_1 = \frac{y}{D^2} \tag{7}$$

Substituting equation 7 in equation 2

$$(N_{FM})^2 = \frac{F_My}{3D^2} \tag{8}$$

Dividing equation 8 by equation 1 gives the ratio of FM to AM noise squared

$$\left(\frac{N_{FM}}{N_{AM}}\right)^2 = \frac{F_My}{3D^2F_My} = \frac{1}{3D^2} \tag{9}$$

The improvement factor for wideband FM systems is thus equal to

$$\text{improvement factor}_{WB} = 10 \log 3D^2 \tag{10}$$

For channelized FM systems see Fig. 8.

Assume a narrow channel centered at frequency, F_m. The AM noise may be designated by

$$(N_{AM})^2 = y\Delta F_m \tag{11}$$

and FM noise by

$$(N_{FM})^2 = y_2\Delta F_m \tag{12}$$

Dividing equation 12 by equation 11 gives

$$\left(\frac{N_{FM}}{N_{AM}}\right)^2 = \frac{y_2\Delta F_m}{y\Delta F_m} = \frac{y_2}{y} \tag{13}$$

Since

$$y = kF_d^2$$

and

$$y_2 = kF_m^2$$

$$\frac{y_2}{y} = \frac{F_m^2}{F_d^2} = \frac{1}{D^2} \tag{14}$$

Therefore the ratio of FM to AM noise is given by

$$\left(\frac{N_{FM}}{N_{AM}}\right)^2 = \frac{F_m^2}{F_d^2} = \frac{1}{D^2} \tag{15}$$

The improvement factor for channelized FM systems is thus equal to

$$\text{improvement factor}_{\text{chan}} = 10 \log D^2 \tag{16}$$

The per-channel noise of an FM system may be obtained simply by calculating the AM noise and applying the improvement factor obtained in the foregoing for any given channel. The emphasis characteristic if used need not be considered, providing the actual deviation of the desired channel is known.

References

1. LOAD RATING THEORY FOR MULTICHANNEL AMPLIFIERS, J. T. Dixon, B. D. Holbrook. *Bell System Technical Journal*, New York, N. Y., Oct. 1939.

2. THE SERRASOID MODULATOR, James R. Day. *Electronics*, New York, N. Y., Oct. 1948.

◆

Discussion

R. S. Caruthers, (Lenkurt Electric Company, Inc., San Carlos, Calif.): Since 1954, when the paper was originally presented, the *72B* radio system has been modified to broaden its range of application to the telephone industry. The modified version has been designated as type *72C*. This discussion compares the *72B* with the *72C* and notes the possible applications of the latter.

The *72B* radio system was designed to provide 72 toll-quality telephone channels for medium-length routes where neither the small-capacity nor the large-capacity systems such as *TD-2* were applicable. The *72C* radio system with 120-channel capacity was the next logical step toward filling the gap between small- and large-system capacity.

In addition to nearly doubling the channel capacity, the *72C* system includes an optional directional-filter arrangement designed to permit three transmitters and three receivers to be operated over the same antenna. Thus, using the frequency assignment shown in Fig. 1, up to 360 voice channels can be transmitted over any multisection system using only two antennas for each section. When 240 or fewer voice channels are required, three radio channels can be installed with one channel acting as stand-by for either of the other two. For a single-section system, two antennas can be used at each terminal and a total of 720 voice channels transmitted over six parallel radio channels using the same route. All six radio channels can be transmitted in the 890 to 940-mc fixed service band.

The industrial band from 952 to 960-mc is wide enough to accommodate one radio channel carrying 120 voice channels.

The increased channel capacity of the *72C* radio system dovetails with the development of new models of type-*45BX* carrier channelizing equipment. This equipment provides up to 120 voice channels in a base band frequency range which is directly applicable to *72C*, from 12 to 528 kc. In addition, the *72C* base band has been extended to 552 kc in order to accommodate a Bell System supergroup (60 channels using frequencies from 312 to 552 kc). This permits the *72C* system to be used with Western Electric type-*L* carrier or with European-manufactured radio carrier systems that also utilize this frequency allocation.

Several minor changes were included in the modification of the *72B* to increase the reliability and improve performance. One of the most important changes was that of improved limiter circuits, which permits achievement of the full FM improvement for signal levels down to threshold. A reduction in the receiver noise figure exceeding 2 db has been achieved, with the use of recent low-noise mixer diodes, together with improved matching and better intermediate-frequency amplifier performance. A modification to the modulator has reduced the basic modulator noise appreciably. Changes have been made in components and in the ventilation, which have improved reliability at high temperature.

Cyclic Decimal Codes for Analogue to Digital Converters

JOSEPH A. O'BRIEN
NONMEMBER AIEE

CYCLIC decimal codes may be used in analogue to digital converters of the reading type. Decimal number output is easily obtained for display to human operators or processing by decimal calculators.[1,2,3] Decimal numbers, of course, are more generally used than binary and more frequently in simple computations.

For definiteness, this paper is limited to a discussion of an encoder in which each measurable angular position of an input shaft is represented by a unique combination of conducting and nonconducting segments. These segments may be on a flat disk having many concentric tracks, as in Fig. 1, or segments may be on a cylinder having many adjacent tracks about the same axis, as in Fig. 2. All conducting segments are connected to a common voltage source. The remarks of this paper are equally applicable, however, to other analogue to digital converters which assign to a physical position a unique code group which may be easily converted to a decimal number form.

A brush in contact with a particular track may be either conducting or nonconducting at a particular time depending on whether it is on a conducting or nonconducting portion of its track. The brushes are then used as 2-state or binary elements. N independent binary elements can assume 2^N different combinations of states. Since $2^3=8$ and $2^4=16$, at least 4 independent binary elements must be used to represent the 10 decimal digits 0 to 9.

Four tracks and their associated brushes are used to assume, in sequence, 10 different combinations of binary states. A particular combination of states is consistently used to represent a particular decimal digit in equipment associated with the encoder. Since the 10 digits could be arbitrarily represented by any 10 and the 16 possible combinations of states of 4 binary elements, $16!/6!$ or more than 29 billion binary decimal codes are possible.[4]

The present discussion is a systematic investigation of codes which will optimize the following parameters:

Readout error.
Number of encoder brushes.
Average brush current.
Number of translator components.

Cyclic Number Sequence and Cyclic Digit Code

Fig. 3 shows an array of conventional decimal numbers with a line passing through successive numbers. When the line moves from left to right, successive numbers differ in only one digit. When the line moves from right to left to connect successive numbers, these numbers differ in more than 1 digit. Fig. 4 shows a cyclic number sequence. The line connecting successive numbers moves from left to right or right to left making a change of only one digit between successive numbers. A cyclic number sequence is defined as a sequence of numbers in which pairs of successive numbers differ in only one digit. The digits which differ are successive digits such as 0 and 1, 1 and 2, 8 and 9, or 9 and 0.

Readout error and the number of brushes may be minimized by using a cyclic number sequence and a cyclic digit code. If a conducting brush is represented by 1 and a nonconducting brush is represented by 0, the states of four brushes at a particular time may be represented by a 4-digit binary number which in turn represents a decimal digit. A cyclic decimal code is defined as a set of 10 words, each containing four binary digits, in which successive words including the first word and the last word differ in only one binary digit.

An angular position encoder which uses both a cyclic number sequence and a cyclic digit code, will have only one brush change from conducting to nonconducting or the reverse in going from one sector to the next. In one complete revolution of the input shaft all consecutive pairs of words read from the brushes differ in only one binary digit. In this respect the code is similar to the reflected binary code and only one set of brushes lying on a radial line is required.

Encoders using a conventional number sequence require at least two brushes per track to prevent errors in readout that might occur at shaft positions where the output number changes in many digits at one time. Brushes, in general, will not lie on radial lines.

Map Method of Code Analysis

The 16 combinations of states of four binary elements may be represented by the map shown in Fig. 5.[5,6] The four successive values of both the horizontal and vertical co-ordinates are 00, 01, 11, and 10. If the vertical co-ordinate is termed *ab*, and the horizontal co-ordinate is termed *cd*, the co-ordinates of a square *abcd*, are a 4-binary digit number. Each square on the map represents a unique 4-digit binary number or word. The decimal number within the square is equal to the binary number expressed by the co-ordinates of that square.

Squares that are geometrically adjacent differ in only one binary digit. Two squares at opposite edges of the map, in the same horizontal row or in the same vertical column, differ in only one binary digit. Squares located symmetrically about a horizontal (or vertical) centerline also differ in only one binary digit. These conditions are defined as "adjacent." Designing a cyclic decimal code is then a simple geometry problem. A progressive sequence of 10 squares must be chosen so that successive squares, including the 10th square and the first square are adjacent.

Two pairs of consecutive digits, 4 and 5, and 9 and 10, are also 9's complement pairs. Each of these four digits must be adjacent on the map to its 9's complement. If the six additional digits are also chosen so that any digit and its complement differ in one particular binary digit, 9's complementing is easily mechanized. The complementing of 9's is required to convert from cyclic to conventional

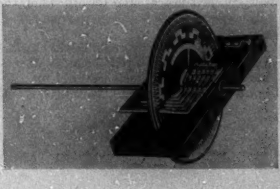


Fig. 1. Angular position disk encoder

Paper 56-21, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted June 23, 1955; made available for printing November 15, 1955.

Joseph A. O'Brien is with the Bell Telephone Laboratories, Inc., Whippany, N. J.

This study was conducted under the sponsorship of the Evans Signal Laboratory, Fort Monmouth, N. J., on Contract *NOa(s)55-869.*

Vol.
Pt. 1
1956

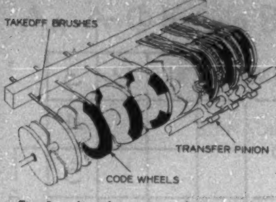


Fig. 2. Angular position drum encoder

AB\CD	00	01	11	10
00	0	1	3	2
01	4	5	7	6
11	12	13	15	14
10	8	9	11	10

Fig. 5 (left). Decimal map

x	x	x	x
x			

CLASS 1

x	x	x	
x	x		

CLASS 2A

x	x	x	
x		x	

CLASS 2B

x	x	x	
	x	x	

CLASS 2C

x	x	x	
		x	x

CLASS 2D

Fig. 6 (right). Basic half maps

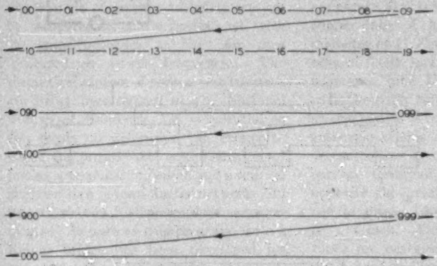


Fig. 3. Conventional number sequence

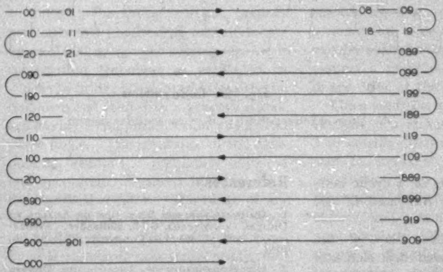


Fig. 4. Cyclic number sequence

AB\CD	00	01	11	10
00	0	1	2	1
01	1	2	3	2
11	2	3	4	3
10	1	2	3	2

Fig. 7. Decimal symmetry map

number sequence. Code analysis is simplified if a digit and its 9's complement occupy squares on the map symmetrical about a horizontal centerline. (Identical results would be obtained from symmetry about a vertical centerline.)

Initial investigation proceeds using half-maps of eight squares, five of which are occupied by digits. Fig. 6 shows the half-maps in which five squares may be connected by a line lying entirely within the map. This classification is rather arbitrary, but gives an illustrative example from the set of 48 maps that must be investigated. Five squares may be chosen from a group of eight squares in ${}_5C_8=8!/5!3!=56$ ways. Of these, ${}_4C_7=7!/4!3!=35$ selections contain the word "0000." Twenty-one selections (${}_5C_7=7!/5!2!$) do not contain the word "0000." Of the 35 selections, 30 meet the requirement of a cyclic decimal code which produces 9's complement by complementing one particular binary digit. Of the 21 selections, 18 form cyclic decimal codes.

Permutation of a code corresponds to relabelling the brushes in a different order or shifting the relative positions of four tracks radially without changing the identifications of the brushes. Complementation of a code corresponds to interchanging conducting and nonconducting segments in a particular track. Complementation of any of the 48 codes produces another code which is a member of the set of 48. Analysis of 48 half-maps then, is a rigorous investigation of all desirable possibilities.

A systematic way of specifying a particular half-map would list the three empty squares with the use of the decimal digits found in the squares of Fig. 5. Thus the map listed in Fig. 6 as class 1 might be called code 567.

Fig. 7 shows a decimal symmetry map, i.e., the number within a square is the sum of the number of 1's in the coordinates of that square. Since a 1 corresponds to a conducting brush, this map is very useful in choosing codes that have words containing the smallest number of 1's.[7] In choosing a binary decimal code to use minimum brush current, it would be desirable to use the following conditions: one word containing four zeros, four words each containing one 1, and five words each containing two 1's. That is, no more than two brushes are ever conducting at the same time. This would give a sum of 14 conducting brushes for the 10 decimal digits. Unfortunately, this code is not cyclic. Fig. 8 shows the code in each class containing the smallest sum of 1's of the coordinates of the occupied squares. Classes 1, 2*a*, and 2*c* (codes 567, 347, and 357) are the only codes having sums of 15. Classes 2*b* and 2*d* (codes 347 and 167) have sums of 17, as do seven additional maps not shown.

Since cyclic codes cannot be easily mechanized for addition or other arithmetic operations, it is desirable to translate to a constant positive weight code. A cyclic decimal code is desirable that can be translated to the 1, 2, 4, 8 code with the minimum number of components. The map of the translator output code is shown in Fig. 9; Capital letters *A*, *B*, *C*, and *D* are used to represent the output lines. If translator input lines are labeled with lower case letters *a*, *b*, *c*, and *d*, the

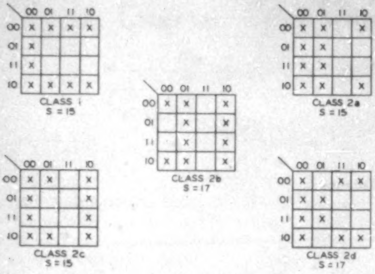


Fig. 8. Minimum brush current codes

Fig. 11 (right). Optimum drum code map: 21 springs

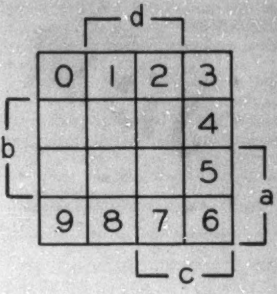
$A = ac'$

$B = c(a+b)$

$C = b'c$

$D = a'(c'd + b'cd') + a(b + c'd' + cd)$

$S = 17$



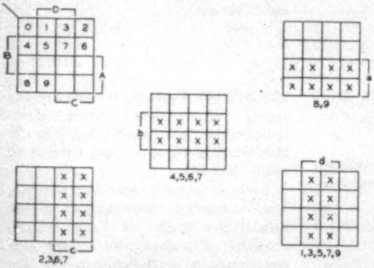


Fig. 9. Binary coded decimal translator output maps

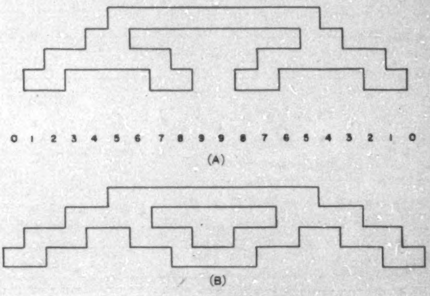


Fig. 12. Code patterns

A—Drum encoder pattern B—Disk encoder pattern

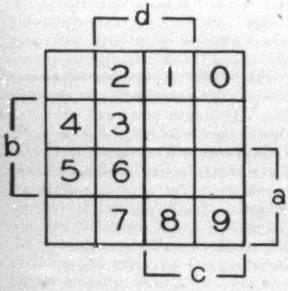


Fig. 10. Optimum disk code map: 23 springs

$A = ac$

$B = c'd' + ac'$

$C = c'd$

$D = a(b'c' + d') + a'd(b + c)$

$S = 19$

squares in which digits of a cyclic code should be chosen are indicated in the remaining maps.

Figs. 10 and 11 show maps of the optimum codes for coded disk elements and coded drum elements respectively. Since a minimum brush current code will not also result in minimum number of translator relay springs and the reverse, the "optimum" codes are the best compromise between these two conflicting requirements.

Here, a disk code is defined as a code that does not use the word "0000," and a drum code does use the word "0000" Fig. 12(*A*) and (*B*) show the patterns these codes will form on the coded elements. In Fig. 12(B) a permutation of the code in Fig. 11 is used to obtain the most desirable placement of tracks.

References

1. Criteria for the Selection of Analog to Digital Converters, **G. L. Hollander.** *Proceedings*, National Electronics Conference, Chicago, Ill., 1953.

2. Interconversion of Analog and Digital Data in System for Measurement and Control, **B. L. Lippel.** *Ibid.*, 1952.

3. A High-Precision Analog-to-Digital Converter, **B. L. Lippel.** *Ibid.*, 1951.

4. Synthesis of Electronic Computing and Control Circuits (book). Harvard Computation Laboratory Series, Harvard University Press, Cambridge, Mass., vol. 27, 1951, chap. XI.

5. The Map Method for Synthesis of Combinational Logic Circuits, **M. Karnaugh.** *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 593–99.

6. A Chart Method for Simplifying Truth Functions, **E. W. Veitch.** *Proceedings*, Association for Computing Machinery, Pittsburgh, Pa., May 1952.

7. The Recognition and Identification of Symmetric Switching Functions, **S. H. Caldwell.** *AIEE Transactions*, vol. 73, pt. I, May 1954, pp. 142–47.

An A-C Kelvin Bridge for the Audio-Frequency Range

BERNADINE L. DUNFEE
ASSOCIATE MEMBER AIEE

THE present trend in electrical measurements indicates the need for a test method by which the ratio and phase angle of current transformers can be measured at adiuo frequencies with the same accuracy as they are now measured at power frequencies. The theoretical aspects of several test methods are being investigated at the National Bureau of Standards. It is essential to this study to determine the extent to which the present test equipment,[1] comprising 4-terminal impedors and a mutual inductor in a potentiometer network, can be utilized in high-frequency measurements. As part of this program, an a-c Kelvin bridge has been developed for measuring the characteristics of the 4-terminal impedors at audio frequencies. At present the bridge method permits resistance and phase-angle measurements up to 6,000 cycles per second (cps). A substitution technique is employed in which the bridge is balanced first with the unknown and then with a primary standard having a computable skin effect and phase angle. The resistance of the computable standard is previously adjusted and accurately measured in terms of a d-c standard resistor. The range of adjustment for the standard extends from 0.1 to 0.05 ohm. With additional computable standards now under study, and minor modifications of the circuit, measurements can probably be extended to a minimum resistance of 0.001 ohm and a maximum frequency of 10^4 cps.

The Computable A-C Standard

A reference standard of a-c resistance must be designed so that its inductance and ratio of a-c to d-c resistance (characterized by the term skin effect) may be computed. A coaxial arrangement with the potential leads attached to either the inner or outer tube is one of the few configurations for which these quantities can be evaluated. The computable standard to be described is of the coaxial type; the arrangement of its potential leads provides a minimum phase angle and maximum protection from coupling between its potential circuit and other current-carrying parts of the network.

CONSTRUCTION

A cross-sectional view of the standard appears in Fig. 1(A). The outer seamless copper tube A has a nominal outside diameter of $1^1/_8$ inches and a wall thickness of 0.065 inch. The inner seamless resistance tube C is arranged coaxially with the outer tube and both are brazed at one end to a connecting ring B. An extension of this ring is necked down to accommodate a coaxial connector. The current terminals, located at the end opposite the brazed joint, are $2^1/_2$-inch square copper blocks having a thickness of 1/4 inch. The ends of the coaxial tubes are centered in these blocks and are brazed to them around their entire circumference. A system of cuts through the blocks, Fig. 1(B), in the form of a symmetrical labyrinth provides four equally spaced entries for the current and assures a more nearly uniform current distribution at the ends of the tubes.

The inner tube, serving as the resistance element, is a nickel-chromium-iron alloy. The outside diameter is 5/8 inch and the wall thickness is 0.01 inch. It has a resistivity of the order of 92 microhm-centimeters (cm) (about 50 times that of copper) and a temperature coefficient of approximately 125 parts per million per degree centigrade. (This relatively high temperature coefficient would be undesirable in most applications. However, with the measuring technique employed here, the resistance of the shunt for a given setting is measured on direct current just prior to its use in the a-c bridge so that the resistance remains constant so long as the current is adjusted to the same value.) For a given configuration, the combination of a thin wall and high resistivity reduces both the phase angle and the skin effect. The over-all length of the standard is about 144 cm with an active length of 133 cm for a 0.1-ohm setting.

A brass rod W, 0.125 inch in diameter and 250 cm long, is located along the axis. At the end where the tubes are brazed together the end extends through an insulating bushing into the constricted portion of B. The end of the rod as well as the cylindrical section of B form the potential terminals P_1 and P_2 and the potential leads are brought away as a coaxial pair with the use of modified *AN* fittings. At the other end, where the current terminals are located, the brass rod emerges from the impedor and extends into the control mechanism along its axis. The control mechanism moves along the rod into cylinder C and, by means of an expandable spider, forms the potential connection p.

The control assembly formed from three coaxial tubes and equipped with suitable adjustments is rigidly attached at one extremity to a vertical arm which rests on a track extending out 110 cm from the impedor. When a clamp that locks the vertical arm to the track is released, the unit can be moved into the resistance tube. The track and the platform on which the impedor is mounted are formed from a single length of brass rigidly fastened to a wood base. The supports for the impedor and the control mechanism are brass, insulated from the metal base, and arranged so that the entire assembly is aligned.

The procedure by which the potential connection p is effected is illustrated in Fig. 2. The inner tube I has an externally threaded section near its end and terminates in a slotted knife-edge taper. The middle tube M carries a fixture threaded both internally and externally and terminates in a double conical wedge K. The outer tube O has an internally threaded section and terminates in a slotted knife-edge taper. Control knob F can be rotated to operate the double wedge. Movement of the wedge into the mating cone expands the fingers of the outer knife edge forcing them to grip the inner wall of the resistance tube. Turning control knob E moves the inner tube I into the wedge causing its fingers to grip the brass rod and completes the operation. A differential screw S coupled to the outer tube O provides a fine adjustment.

SKIN EFFECT

The resistance of the a-c standard must be measured eventually on direct current. It is necessary, therefore, to design the shunt so that its skin effect is either negligible or computable. The skin effect of conductors and shunts has been treated by a number of writers.[2] The effect in

Paper 56-25, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 17, 1955; made available for printing November 10, 1955.

BERNADINE L. DUNFEE is with the National Bureau of Standards, Washington, D. C.

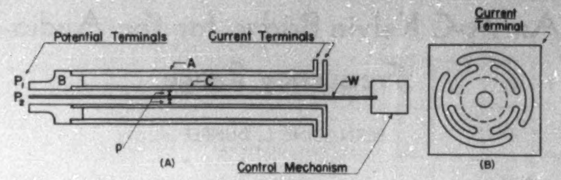


Fig. 1(A). Cross section of the computable standard. (B). Current terminal block

tubular-type shunts, completely defined by the four parameters, namely, resistivity ρ, permeability μ, wall thickness d, and frequency f can be computed from an equation developed by Silsbee.[2] At the operating frequencies of the computable standard, the a-c resistance is given closely enough by the equation

$$R_{ac}=R_{dc}\left(1-\frac{7m^4d^4}{360}\right)$$

where $m^2=8\pi^2 f/\rho$. The permeability is assumed equal to unity. The ratio of its a-c to d-c resistance at 10^4 cps is $(1-4\times 10^{-7})$.

Phase Angle from Geometrical Considerations

Silsbee has made an extensive analysis of this problem for several configurations.[2,4] In his treatment of coaxial assemblies, expressions for the regional flux linkages are obtained from the circuital form of Ampere's law. The inductance for a given configuration is given by the ratio of the active flux linkages Φ to the current I where Φ is the sum of only those linkages instrumental in producing an electromotive force (emf) across the potential terminals. The corresponding phase angle is for an impedor used in a potentiometer network where, at balance, there is no current in the detector circuit and, therefore, none in the potential leads of the shunt.

The expression that defines the inductance is altered when the impedor becomes an element of the Kelvin bridge network. The small currents in the potential leads of the impedor become sources of magnetic flux and introduce additional self- and mutual-inductance terms. To interpret correctly the influence of these lead currents as dictated by the bridge balance equation, it was necessary to extend Silsbee's analysis and to obtain the expression for the inductance, taking into account the altered current distribution. Computed in this manner, the algebraic sum of the active flux linkages per unit length in a standard of the given configuration is given by

$$\Phi=\Phi_{CIC}-\Phi_{CIW}-\Phi_{EWW}-\Phi_{WWW}$$

Referring to the view of Fig. 3(A) as seen from a cut across the diameter, these flux linkages are identified as follows:

Φ_{CIC} = the average flux linkages in region C arising from the total current I entering the shunt. This current, Fig. 3(B), includes the current in the rod W designated as δi, and that in the resistance tube designated as $I-\delta i$

Φ_{CIW} = the flux in region C, produced by the current I ($I-\delta i$ plus δi), linking the rod W

Φ_{EWW} = the flux in region E, produced by the current δi in the rod, linking the rod W

Φ_{WWW} = the average flux linkages in region W produced by the current δi in the rod

The linkages expressed in terms of the currents and the dimensions of the inner tube and rod are (see the Appendix for a typical derivation)

$$\Phi_{CIC}=\left(t+\frac{t^2}{6}-\frac{t^4}{30}\cdots\right)\delta i+\left(\frac{2}{3}t-\frac{t^3}{15}-\frac{t^4}{20}\cdots\right)(I-\delta i)$$

$$\Phi_{CIW}=\left(2t+t^2+\frac{2}{3}t^3\cdots\right)\delta i+\left(t+\frac{t^2}{6}-\frac{t^4}{30}\cdots\right)(I-\delta i)$$

$$\Phi_{EWW}=\left(2v+v^2+\frac{2}{3}v^3\cdots\right)\delta i$$

$$\Phi_{WWW}=\frac{1}{2}\delta i$$

where

$t=(a_2-a_1)/a_2$
$v=(a_1-r)/a_1$
a_2 and a_1 = the outside and inside radii of the resistance tube C
r = radius of the rod W

By definition the series expansion terms express the respective inductances in the centimeter-gram-second system of units. It will be noted that Φ_{CIC} as well as Φ_{CIW} has two components: flux linkages associated with a self-inductance L and those associated with a mutual inductance M. These can be written, therefore, in the forms

$$\Phi_{CIC}=M_{CW}(\delta i)+L_C(I-\delta i)$$

$$\Phi_{CIW}=L_W(\delta i)+M_{WC}(I-\delta i)$$

The flux linkages $\Phi_{EWW}=L_W'\delta i$ and $\Phi_{WWW}=L_W''\delta i$ are associated with self-inductances. The question arises regarding the significance of these components and their proper selection in computing the phase angle of the standard. The self-inductances L_W, L_W', and L_W'' of the rod are lead inductances which, like the lead resistance of the d-c network, need be considered only in the examination of the auxiliary lead balances. It can be shown that the phase angle of the standard must include only the components L_C and M_{WC} if the phase angle of the unknown is to be valid for a potentiometer network when computed from the bridge equation. The remaining quantity M_{CW} may be considered as an extraneous mutual inductance since its presence gives rise to an induced emf that is registered in the bridge circuit. Its effect, if appreciable, enters the balance equation as a correction term. The phase angle of the standard was computed as 3.1 minutes at 10^4 cps and the effect from the extraneous mutual coupling was found to be negligible.

The A-C Kelvin Bridge[5,6]

The D-C to A-C Transformation

Referring to Fig. 4, the balance equation for the d-c network is[7]

$$\frac{A}{B}=\frac{E}{F}\left(\frac{1+\alpha/E}{1+\eta/F}\right)+\frac{L(D+\gamma)}{B[(C+\beta)+L+(D+\gamma)]}\left(\frac{E+\alpha}{F+\eta}-\frac{C+\beta}{D+\gamma}\right)\tag{1}$$

The two auxiliary balance equations are

$$\frac{A}{B}=\frac{\alpha}{\eta}+\frac{L(D+\gamma)}{B[(C+\beta)+L+(D+\gamma)]}\left(\frac{\alpha}{\eta}-\frac{C+\beta}{D+\gamma}\right)\tag{2}$$

$$\frac{A}{B}=\frac{E+\alpha}{F+\eta}+\frac{D+\gamma}{B}\left(\frac{E+\alpha}{F+\eta}-\frac{C+\beta}{D+\gamma}\right)\tag{3}$$

where equation 1 is the balance equation for the main bridge shown in Fig. 4, equation 2 is the balance equation for the auxiliary bridge formed by short-circuiting units E and F with a link L' of negligible resistance, and equation 3 is the balance equation for the auxiliary bridge formed when links L and L' are both removed. The balances indicated

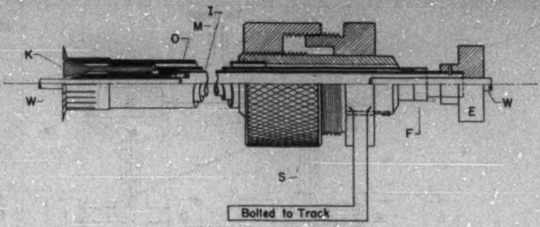


Fig. 2. Control mechanism

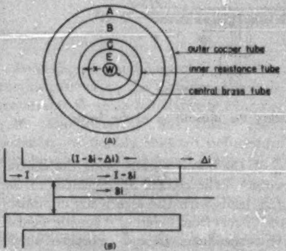


Fig. 3(A). Regions used in flux linkage calculations. (B). Current distribution

by equations 2 and 3 permit adjustment of the lead resistances α, β, γ, and η, including the resistances of the potential leads l_1, l_2, l_3, and l_4, into unique ratios such that equation 1 can be reduced to the simple form

$$\frac{A}{B}=\frac{E}{F}$$

With resistances replaced by impedances, the balance equations for the a-c network are similar in form to those for d-c operation. It is to be noted, however, that the link resistance L which appears in the numerators of equations 1 and 2 becomes an impedance of the form $Z_L' = Z_L + j\omega(M_1 - M_2)$ in the a-c case and includes, in addition to the impedance Z_L, associated with the link, the inductive reactance associated with these emfs of mutual inductance developed in the potential leads of A and B common to loop III. If the emf of mutual inductance developed in the 4-terminal resistors A and B by the currents in their potential leads (l_1, l_2, l_3, and l_4) are neglected, the phase angle as well as the resistance of resistor A, or B, given by these equations is the one to be measured, i.e., the phase angle of the resistor in a potentiometer network. (However, the inductive reactance terms associated with the emf of mutual inductance developed in the potential leads by currents in the current circuits of the shunts are included in the appropriate impedance term of equations 1, 2, or 3.)

Inclusion of these emf of mutual inductance, however, alters the equation; typically $Z_A/Z_B = \zeta\,(1+a+jb)$ where ζ represents the right-hand side of equation 1, 2, or 3 in its a-c form. Whether or not the correction $(a+jb)$ is negligible depends on the magnitudes of the bridge elements and on the mutual inductances present in the 4-terminal shunts. In the substitution method, only the mutual inductance of the standard and of the unknown is significant since shunt B is common. As previously mentioned, the extraneous mutual inductance is computable for the standard; it is not computable for the unknown so that, to some extent, the correction term is indeterminate. However, when the resistances of the bridge elements are large, the correction term is sufficiently small so that, with some knowledge of the construction of the unknown, its mutual inductance can be estimated with sufficient accuracy. For example, in the measurements of the 4-terminal shunts used in current-transformer testing, the errors introduced by neglecting the term $(a+jb)$, for an assumed mutual inductance of 0.4 microhenry, are less than 3 parts per million for the resistance and 0.1 minute for the phase angle at a frequency of 10^4 cps.

Lead Balances

In the d-c bridge the influence from lead and contact resistances is minimized through auxiliary balances that force these resistances into appropriate ratios. In the a-c bridge, lead inductances as well as resistances must be considered. Assuming that the lead resistances are adjusted to their proper ratios, an analysis shows that the necessity for lead inductance balances can be avoided by making the leads short, operating with elements of large impedance and utilizing the substitution technique. For example, with 1,000 ohms in each of the bridge ratio arms and with an uncompensated lead inductance of 1 microhenry, the resulting error in phase angle would be less than 0.2 minute at 10^4 cps. A switching arrangement provides for a quick transfer of the network from a d-c to an a-c source. Thus, with the network energized on direct current, the necessary lead resistance balances are made. The network is then transferred without change to an a-c source and the main bridge is balanced, no lead inductance balances being required.

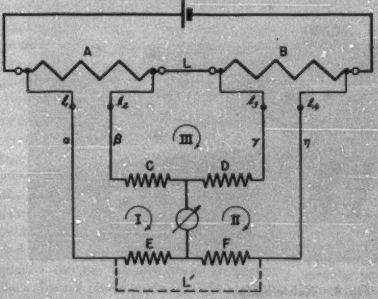


Fig. 4 (left). The d-c Kelvin bridge network

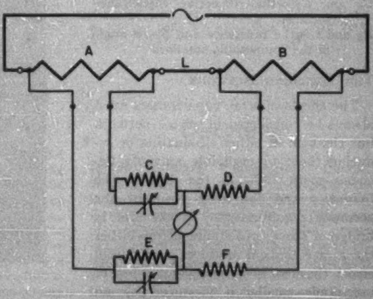


Fig. 5 (right). The a-c bridge without the shielding network

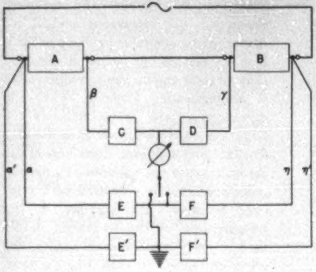


Fig. 6 (above). Circuit indicating Wagner guard

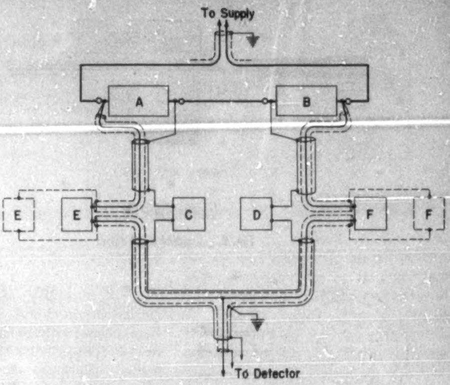


Fig. 7. Circuit indicating the disposition of the leads

Bridge Components and the Defining Equation

Capacitors rather than inductors were chosen as the reactive bridge balance elements to avoid the spurious flux linkages that characterize an inductive network. The bridge circuit without its shielding network is shown in Fig. 5. The computable standard (or the unknown resistor) occupies the *A* arm. *B* is a 4-terminal resistor the phase angle of which is more inductive than either the standard or the unknown; *D* and *F* are 1,000-ohm audio-frequency resistance cards. The resistors of arms *C* and *E* are audio-frequency decade resistance boxes, with settings greater than 500 ohms in all cases; the capacitors are continuously adjustable from 100 to 1,100 microfarads. The defining equations for the a-c resistance and the phase angle, when the substitution method is employed, are

$$R_X=\frac{R_E R_S}{R_E'}[1+\omega R_E C_E(\omega R_E' C_E'-\omega R_E C_E)] \tag{4}$$

$$\theta_X \approx (\omega R_E' C_E'-\omega R_E C_E)+\theta_S \tag{5}$$

where

R_E and R_E' = the resistance settings, at balance, of the *E* arm when the unknown and the standard respectively occupy the *A* arm

C_E and C_E' = the respective capacitance settings

R_S and θ_S = the resistance and phase angle of the computable standard

The Shielding Network

The effects of stray capacitances must always be considered in an a-c network but the task of either eliminating or reducing them to negligible amounts becomes more difficult the higher the frequency and the more stringent the accuracy requirements. In the Kelvin bridge, a stray capacitance across either the *E* or *F* element produces a maximum error; the error becomes several orders of magnitude smaller if produced by an equal capacitance across the *A*, *B*, *C*, or *D* elements. A study of this problem as well as pertinent experimental measurements indicated that an adjustable Wagner network would provide the most effective shielding, particularly if its leads α' and η', Fig. 6, were arranged to protect the corresponding leads α and η of the measuring circuit.

The diagrams of Figs. 6 and 7 show the bridge circuit (supply and detector circuits are omitted) and illustrate the disposition of the leads. The α' and η' leads of the Wagner arms are flexible braided shields that protect the critical α and η leads. The β and γ leads of the inner bridge arms form a sheath around the respective coaxial pairs α, α' and η, η', the α and η leads being the central wires. (The coaxial arrangement of the β, γ leads with the α, η leads minimizes the effects from external magnetic fields.) Thus the inner wire and its braided shield together with the outside flexible sheath form a triple coaxial configuration. Special adjustable resistors for balancing the lead resistances were incorporated in the triple coaxial arrangement. Further precautions were taken by connecting to ground each enclosure in which the elements were housed, the physical ground being located at the junction of the Wagner arms. With the capacitances between the elements thus fixed their effects were eliminated by the substitution method.

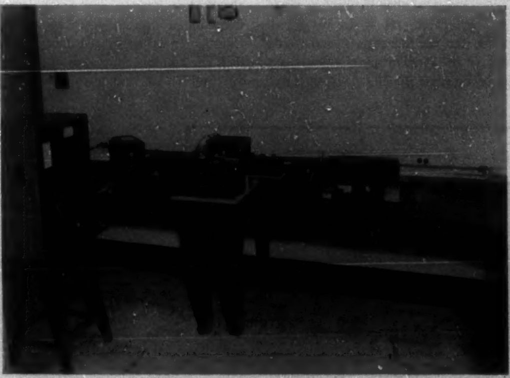


Fig. 8. Kelvin bridge

Vol. 7
Pt. 1
1956

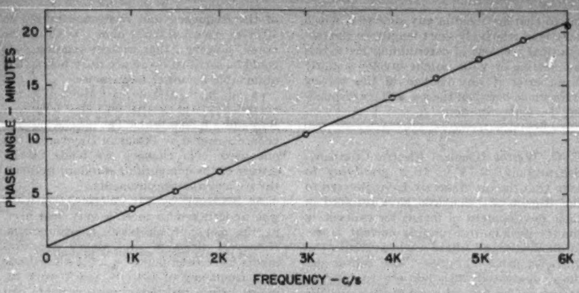


Fig. 9. Plot of experimental data obtained for a 0.1-ohm shunt

To realize the high accuracy desired in the measurement, the detector circuit is required to sense an unbalance of 1 microvolt. The detector has three components: a doubly shielded impedance matching transformer having a 1 to 4 ratio; a preamplifier with a gain of about 30; and a wave analyzer that serves both as a sharply tuned filter and as an indicator of unbalance. The shields of the three units together with their common lead are connected to ground.

Precautions were taken to protect the circuit against external magnetic fields as well as minimize the magnetic coupling between the bridge members. The coaxial arrangement of the inner and outer arms afforded some protection. Loop areas were further reduced by arranging the units in two tiers so that the circuit could, in effect, fold back on itself. The bridge compactness is shown in Fig. 8.

Results

The results obtained in a series of measurements on a 0.1-ohm shunt indicate the usefulness of the test method. The measurements include a-c determinations over a frequency range of 1,000 to 6,000 cps taken on several days over a span of 4 months. Minor modifications, rearrangement of the bridge and supply circuits, and substitution of the bridge elements were made between tests. The modifications added to the convenience of operation but neither they nor the other changes produced appreciable differences in the results. All values of resistance obtained at a given frequency were in substantial agreement and their average was within 20 parts per million of the value measured in the d-c bridge normally used in certification work. No significant skin effect was evident over this frequency range. At any given frequency the values of phase angle agreed to within 0.5 minute. The averages plotted against frequency are indicated in Fig. 9. It will be noted that the experimental values fall on a straight line passing through the origin, justifying the approximations made in deriving equation 5. Furthermore, an extrapolated 60-cycle value is within 0.1 minute of the value obtained by another method in 1931 when the shunt was constructed.

Summary

An adjustable 4-terminal resistor, having a computable phase angle, is described and its a-c performance is analyzed. A shielded a-c Kelvin bridge network suitable for measuring the resistances and phase angles of 4-terminal resistors in the audio-frequency range has been described. Both a-c and d-c sources are available and the network is easily transferred from one to the other. The resistance of the adjustable standard is set to the nominal value of the unknown and measured on direct current in terms of a standard d-c resistor. The bridge is then balanced on alternating current first with the standard and then with the unknown in the network, with lead resistance balances made on direct current prior to the a-c measurements. The validity of the test method is demonstrated by a typical example.

Appendix. Derivation of the Flux Linkages Φ_{CIC}

Let the current in the wire W be δi and that in the cylinder C, $(I-\delta i)$, both having a direction into the paper. Assume the permeability equal to unity and let the inner and outer radii of C be a_1 and a_2. The flux density B at a distance x from the common center is by Ampere's law $\int Bdl = 4\pi I_n$ where I_n is the current enclosed by a circular path of length l having a radius x. Since the current δi in the wire is completely enclosed, and the fraction $(x^2-a_1^2)/(a_2^2-a_1^2)$ of the current in the cylinder $(I-\delta i)$ is enclosed, the foregoing equation when integrated is

$$B=\frac{2}{x}\delta i+\frac{2}{x}\left(\frac{x^2-a_1^2}{a_2^2-a_1^2}\right)(I-\delta i)$$

The flux linkages averaged across region C may be written

$$\Phi_{CIC}=\frac{1}{\pi(a_2^2-a_1^2)}\int_{a_1}^{a_2}2\pi x dx\int_x^{a_2}Bdx$$

$$=\frac{1}{\pi(a_2^2-a_1^2)}\int_{a_1}^{a_2}2\pi x dx\int_x^{a_2}\left[\frac{2}{x}\delta i+\frac{2}{x}\left(\frac{x^2-a_1^2}{a_2^2-a_1^2}\right)(I-\delta i)\right]dx$$

This yields

$$\Phi_{CIC}=\left(-\frac{2a_1^2}{a_2^2-a_1^2}\ln\frac{a_2}{a_1}+1\right)\delta i+\left(\frac{a_2^2-3a_1^2}{2(a_2^2-a_1^2)}+\frac{2a_1^4}{(a_2^2-a_1^2)^2}\ln\frac{a_2}{a_1}\right)(I-\delta i)$$

Define $t=(a_2-a_1)/a_2$. After expanding the logarithmic terms and substituting the value of t as indicated, the equation reduces to

$$\Phi_{CIC}=\left(t+\frac{t^2}{6}-\frac{t^4}{30}\cdots\right)\delta i+\left(\frac{2}{3}t-\frac{t^2}{15}-\frac{t^4}{20}\cdots\right)(I-\delta i)$$

References

1. EQUIPMENT FOR TESTING CURRENT TRANSFORMERS, **Francis B. Silsbee, Ray L. Smith, Nyna L. Forman, John H. Park.** *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 11, July 1933, pp. 93–122.

2. NOTES ON THE DESIGN OF FOUR-TERMINAL RESISTANCE STANDARDS FOR ALTERNATING CURRENT, **Francis B. Silsbee.** *Ibid.*, vol. 4, Jan. 1930, pp. 73–107 (see references).

3. SHUNTS AND INDUCTORS FOR SURGE-CURRENT MEASUREMENTS, **John H. Park.** *Ibid.*, vol. 39, no. 3, 1947, pp. 191–212.

4. A STUDY OF THE INDUCTANCE OF FOUR-TERMINAL RESISTANCE STANDARDS, **Francis B. Silsbee.** *Bulletin, Ibid.*, vol. 13, no. 3, 1916, pp. 375–422.

5. THE FOUR TERMINAL CONDUCTOR AND THE THOMSON BRIDGE, **Frank Wenner.** *Ibid.*, vol. 8, no. 3, 1912, pp. 559–610.

6. MEASUREMENT OF THE INDUCTANCES OF FOUR-TERMINAL RESISTANCE STANDARDS, **L. Hartshorn.** *Proceedings*, Physical Society of London, London, England, vol. 39, 1927, pp. 377–87.

7. ELECTRICAL MEASUREMENTS (book), **Forest K. Harris.** John Wiley and Sons, Inc., New York, N. Y., 1952.

◆

Discussion

H. N. Hayward (University of Illinois, Urbana, Ill.): This paper is of particular interest because of the application made of a Kelvin bridge to measure the a-c resistance of low-valued resistors. It also points out some of the problems that arise in attempting to extend the methods of making accurate measurements from the power frequencies into the audio-frequency range.

The measurements reported were based on the use of a computable a-c standard of resistance having one potential terminal that could be adjusted in position to vary the amount of resistance between the potential terminals. Although not stated in the paper, it is assumed that the procedure used was to set this computable and adjustable standard to the value required in making the a-c measurement and then, without changing the setting of this computable standard, to determine the resistance value of the standard in a d-c bridge. It seems advisable to emphasize this procedure, which would be necessary because of the relatively large error that might result if any setting of the adjustable computable standard should be changed between an a-c measurement and the corresponding d-c determination. This problem is similar to that involved in any slidewire, which here is of relatively short length, so the mechanical problem of determining the actual a-c setting or value might involve a fairly large error if any change in the setting were made between the a-c and corresponding d-c determinations.

C. O. Werres (General Electric Company, Schenectady, N. Y.): It is gratifying to note that the questions we have directed to the Bureau of Standards have resulted in their development of means for calibrating shunts used in determining current transformer ratio and phase angle at audio frequencies above 60 cycles. The author and those associated with her are to be congratulated for this advance of the art. We look forward to the contemplated extension of the frequency and resistance ranges to 10,000 cycles and 0.001 ohm. It should be noted, however, that present requirements could be satisfied by an accuracy below that obtainable at power frequencies.

Bernadine L. Dunfee: The measuring procedure assumed by Professor Hayward is the one used. No changes are made in the setting of the computable standard between the a-c and d-c measurements.

As indicated by Mr. Werres, this work was undertaken to meet a very real need in the field of electrical measurements. However, although preliminary measurements have been made on a 0.1-ohm shunt at a frequency of 10^4 cps, much work remains before resistance measurements can be extended to 0.001 ohm at this frequency.

A First Look at Random Noise

S. O. RICE
NONMEMBER AIEE

Synopsis: The aim of this paper is to discuss some elementary properties of a type of noise which has been called "Gaussian" or "random." The power spectrum is discussed and its relation to the autocorrelation function is pointed out. Two probability distributions associated with the noise are described.

OCCASIONALLY one encounters engineering problems which involve the random fluctuations of an electric current or voltage, or some other quantity. These fluctuations are often called "noise." A considerable amount of information about such fluctuations has been obtained by mathematicians who have studied the subject under the name of "stochastic processes." The aim of this paper is to present an introduction to the subject by mentioning some of the main properties of the fluctuations, which in accordance with engineering usage, will be spoken of as noise.

There is a wide variety of noises, but this paper will be restricted to those called "stationary processes." That is, they go on about the same forever, and their statistical properties do not change with time. Furthermore, the discussion will be concerned mostly with a special kind of noise which has been called "random" or "Gaussian" noise. Its second name comes from the fact that its statistical properties are closely related to the Gaussian distribution used in probability theory. Random noise is a mathematical idealization of certain types of noise occurring in nature. One such type is thermal noise, which occurs as a result of the thermal agitation of electrons in a conductor.

Noise consisting of impulses that can be heard separately is not considered to be random noise in the sense that the term will be used here. For example, noise from relay clicks and ignition noise from automobile engines are not random noises. When such trains of impulses are expressed as Fourier series or as Fourier integrals, there is a considerable amount of correlation between the phases of the various components. This is not true for random noise.

Why are engineers interested in the properties of noise? One reason is that noise produces interference in transmission systems. A knowledge of the statistics of the noise is helpful in estimating the quality of a proposed system while it is still in the design stage. In some cases the noise is large in comparison with the signal, as in the problem of detecting airplanes or submarines. The question here is how best to overcome the noise. Again, carrier communication systems are coming into use which employ a wide band of frequencies to transmit a large number of telephone channels. The output of such a system resembles random noise, and this fact may be used in designing the system.

Random noise has been studied more extensively than other types of noise. The probable reason for this is that it is easier to handle mathematically than the other types. Furthermore, some frequently encountered noises, such as thermal noise in conductors and shot-noise in vacuum tubes, are close approximations to random noise.

An oscillogram of a random noise current might look somewhat like that shown in Fig. 1. Here $I(t)$ is plotted vertically and the time t goes on horizontally. Throughout this discussion, $I(t)$ will stand for a random noise current. The division of the time scale into strips of length T will be needed later in connection with representations for the noise.

What are the various statistical properties of this curve, i.e., such things as the mean square value of $I(t)$, the percentage of time $I(t)$ exceeds some given value I, or the expected number of zeros per second? Some of these may be readily obtained by working directly with the oscillogram. For example, the expected number of zeros per second may be obtained by a count of the number of times the curve crosses the axis during a long interval of time. Other properties may require some calculation; an example is the mean square value of $I(t)$. However, in most problems an oscillogram is not provided. Generally, all that is known about the noise is its power spectrum, about which more will be said later. It may be shown that the statistics of $I(t)$ are completely determined by the power spectrum.

Paper 56-73, recommended by the AIEE Communication Theory Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing January 12, 1956.

S. O. Rice is with Bell Telephone Laboratories, Inc., New York, N. Y.

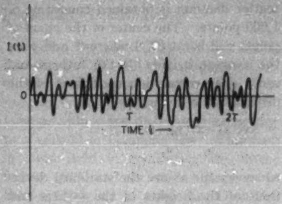
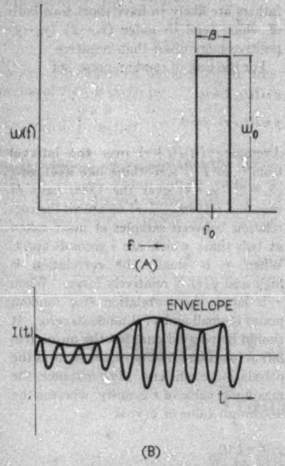


Fig. 1. Portion of a long oscillogram of a random or Gaussian noise current I(t)



(A)

(B)

Fig. 2. Narrow-band power spectrum and corresponding random noise current

Representation of Random Noise

One method often used in the study of random noise is first to represent $I(t)$ by some kind of mathematical expression. There are several representations in use. They take the form of functions of time which contain a number (generally infinite) of random variables as parameters. The random parameters are needed to get the randomness into the noise.

One of the oldest and most frequently used representations is the Fourier series[1]

$$I(t)=\sum_{n=1}^{\infty}\left(a_n\cos\frac{2\pi nt}{T}+b_n\sin\frac{2\pi nt}{T}\right) \qquad (1)$$

where T is a very long interval of time, as indicated in Fig. 1, and the a_n's and b_n's are independent random variables. The representation is so named because it is suggested by cutting the oscillogram of Fig. 1 into strips of length T and expanding $I(t)$ on each strip as a Fourier series. When this is done it is found that the numerical values of the coefficients for any one strip are so irregular that they show no relation to each other (aside from being of the same order of magnitude) or to the corresponding coefficients for the other strips. This is in accordance with the assumption that the a_n's and b_n's are independent random variables. The standard deviations of the a_n's and b_n's are fixed by the power spectrum of $I(t)$.

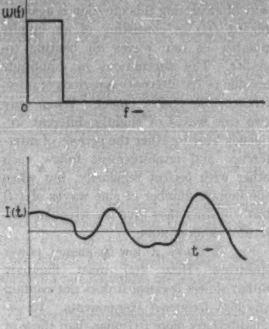


Fig. 3. Low-frequency power spectrum

Representations may be used directly in practical investigations. For example, suppose a random noise voltage is applied to a filter network. What are the statistical properties of the output current? A Fourier series representation for the applied voltage may be written down and the response to each sinusoidal component computed from network theory. The output current is then given by the sum of these response. The result is a Fourier series representation for the current from which the desired statistical properties may be computed.

Again, suppose that a random noise current passes through a hot-wire ammeter. If the noise is fluctuating slowly enough the needle of the ammeter will also fluctuate a little. The rate and the amount of fluctuation will depend upon both the time constant of the meter and the properties of the noise. For example, how can the rms value of the needle fluctuations be calculated? This may be done by starting with the expression

$$D(t_0)=\int_{-\infty}^{t_0} e^{-\alpha(t_0-t)}I^2(t)\,dt$$

It may be shown that the needle deflection at the instant t_0 is approximately proportional to $D(t_0)$. In this integral α depends upon the time constant of the ammeter. When $I(t)$ is replaced by its Fourier series representation, equation 1, and the integration carried out an expression is obtained for $D(t_0)$. The rms value of the needle fluctuations may be obtained by working with this expression.

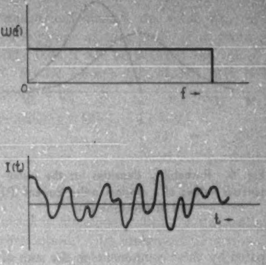


Fig. 4. Wide-band power spectrum

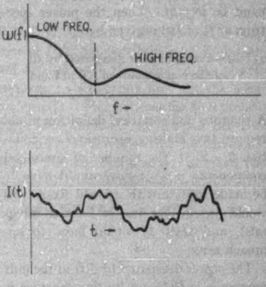


Fig. 5. General power spectrum

The Power Spectrum

As its name indicates, the power spectrum measures the distribution of the power in a noise over the frequency range from zero to infinity. Many kinds of noise besides random noise have power spectra. However, for the sake of simplicity, only the random noise current $I(t)$ will be considered.

Suppose $I(t)$ to be composed of a great many sinusoidal components as indicated by the Fourier series representation, equation 1. Suppose further that $I(t)$ flows through a resistance of 1 ohm. Each component will dissipate its share of power, and the total power dissipated is the sum of the powers dissipated by the components.

Choose some narrow frequency band, say the band between f and $f+\Delta f$. A certain number of components will lie in this band, and the total power dissi-

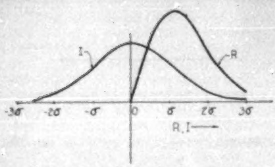


Fig. 6. Probability densities for the instantaneous values of a random noise current I(t) and its envelope R(t)

pated by these components is the sum of the powers of the individual components. Let the average of the total power in the band be $w(f)\Delta f$. Then the power spectrum $w(f)$ of $I(t)$ is given by

$$w(f)\Delta f = \text{average power dissipated by all of the components of } I(t) \text{ in band } (f, f+\Delta f) \quad (2)$$

A rigorous mathematical definition would require two limiting operations on equation 2. First, the number of sinusoidal components in $(f, f+\Delta f)$ would have to be made infinite with Δf held fixed (this would require the use of Fourier integrals), and then Δf would have to approach zero.

The power dissipated by $I(t)$ in the unit resistance is $I^2(t)$ and its time average is $\overline{I^2(t)}$. This average power is also equal to the sum of the average powers of all of the components. The sum is obtained by integrating the power spectrum over the entire frequency range. Thus,

$$\overline{I^2(t)} = \int_0^\infty w(f)df$$

When the power spectrum of $I(t)$ is known, something may be said about the general appearance of $I(t)$ as a function of time. Figs. 2-5 show some typical examples of random noise currents and the corresponding power spectra.

Fig. 2 shows a narrow-band power spectrum centered on frequency f_0 and extending from $f_0-\beta/2$ to $f_0+\beta/2$. The first thing the power spectrum shows about $I(t)$ is its rms value. Since $\overline{I^2(t)}$ is given by the area under the $w(f)$ curve, and since this area is $w_0\beta$, then

$$\text{rms } I(t) = \sigma = (w_0\beta)^{1/2}$$

Also, since the power spectrum is concentrated around the mid-band frequency f_0, it indicates that $I(t)$ will behave much like a sine wave of frequency f_0 with a slowly fluctuating amplitude or envelope. This behavior is shown in Fig. 2. The narrower the band, the more slowly the envelope fluctuates. In fact it may be shown that the envelope may be expected to have about 0.64 β maxima per second.

Why does the envelope fluctuate? One way of answering this question is again to regard $I(t)$ as consisting of an infinite number of sine waves all beating together. The fluctuations in amplitude are due to the alternating periods of interference and reinforcement. When only two sine waves of slightly different frequency beat together the periods of interference and reinforcement follow each other with perfect regularity, but when the infinite number of sine waves of $I(t)$ beat together they give the irregular pattern of random noise.

Fig. 3 shows a low-frequency power spectrum. Here $I(t)$ must fluctuate rather slowly because it does not contain any high-frequency components. Its oscillations are more irregular than those of narrow-band random noise because the "envelope" is fluctuating almost as fast as is the wave itself. Fig. 4 shows a wide-band power spectrum which extends down to zero frequency. Here the fluctuations of $I(t)$ are much more rapid than in the case of Fig. 3, but the degree of irregularity is about the same. In fact, the two cases are really the same statistically except for a change in the time scale.

Fig. 5 shows a general type of power spectrum which may be roughly divided up into two parts; a low-frequency portion and a high-frequency portion. The low-frequency portion gives the slow, large-amplitude fluctuations of $I(t)$ while the high-frequency portion contributes the rapid, small fluctuations which are superposed on the slower variations.

It is hoped that the curves drawn for Figs. 2, 3, 4, and 5 are true enough to life to give an idea of how the appearance of a random noise current is affected by the shape of its power spectrum.

The Autocorrelation Function

An important function associated with a given noise current is its autcorrelation function. The autcorrelation function $\psi(\tau)$, is defined as the time average of the product $I(t)I(t+\tau)$ and is a function of τ.

$$\psi(\tau) = \overline{I(t)I(t+\tau)} = \underset{T\to\infty}{\text{Limit}} \frac{1}{T}\int_0^T I(t)I(t+\tau)dt$$

$\psi(\tau)$ is much like the correlation coefficient in statistics. One of the classical examples used to illustrate the correlation coefficient is the relation between the heights of fathers and sons. Suppose measurements are obtained of the heights of 1,000 men and their fathers. If y_n is taken to be the height of the nth son and x_n the height of his father, and the measurements plotted on the (x,y) plane, a scatter diagram is obtained consisting of 1,000 points. The center of the cloud of points will be at $(\bar{x},\bar{y})$ where $\bar{x}$ and $\bar{y}$ are the average heights of the fathers and sons. The correlation coefficient of the distribution is defined as

$$r = \frac{\text{Ave}(x_n-\bar{x})(y_n-\bar{y})}{\sigma_x \quad \sigma_y}$$

where σ_x and σ_y are the standard deviations of the heights of the fathers and sons. In this particular example, the coefficient r is positive because tall fathers are likely to have tall sons and short fathers are likely to have short sons both of which tend to make $(x_n-\bar{x})\ (y_n-\bar{y})$ positive more often than negative.

For the case of random noise, set

$$x = I(t),\ \bar{x} = 0,$$

$$y = I(t+\tau),\ \bar{y} = 0$$

Averaging $I(t)I(t+\tau)$ over the interval from 0 to T is something like averaging $(x_n-\bar{x})\ (y_n-\bar{y})$ over the 1,000 pairs of sons and fathers. $\psi(\tau)$ measures the correlation between samples of noise taken at two times which are τ seconds apart. When τ is small, the correlation is high and $\psi(\tau)$ is relatively large. When τ is large, the correlation (for random noise) is small and $\psi(\tau)$ tends to zero. It should be pointed out that the autocorrelation function is not quite the same as the correlation coefficient. For instance, the maximum value of r is unity, whereas the maximum value of $\psi(\tau)$ is

$$\psi(0) = \overline{I^2(t)}$$

There is a remarkable relation between the power spectrum and the autocorrelation function, namely, one is the Fourier cosine transform of the other:

$$\psi(\tau) = \int_0^\infty w(f) \cos 2\pi f\tau\, df$$

$$w(f) = 4\int_0^\infty \psi(\tau) \cos 2\pi f\tau\, d\tau$$

When either $w(f)$ or $\psi(\tau)$ is known, then the other one can be obtained—by numerical integration if need be. These equations are sometimes called the "Wiener-Khintchine" relations. They are of considerable importance in solving practical problems. For example, it is often of interest to compute the power spectrum of the noise appearing in the output of some nonlinear device, such as a detector or amplifier, when the input consists of a signal plus noise. In such problems it is usually much easier to compute the autocorrelation function first and then obtain the power spectrum by integration.

Probability Distributions

Finally, the probability distribution of the values of $I(t)$ will be considered. One of the most important properties of a random noise current is that its instantaneous values are distributed according to a normal or Gaussian law. This fact is often used in computing the interference produced by noise. Suppose a time t_1 is chosen at random. Then the probability that $I(t_1)$ will lie between I and $I+dI$ may be shown to be

$$\text{Probability } [I<I(t_1)<I+dI] = \frac{dI}{\sigma\sqrt{2\pi}} \exp\left[-\frac{I^2}{2\sigma^2}\right]$$

where the standard deviation (the rms value of $I(t)$) is given by

$$\sigma^2 = \overline{I^2(t)} = \int_0^\infty w(f)df$$

It is important to note that the distribution of the instantaneous values of $I(t)$ does not depend on the shape of $w(f)$. The distribution depends only on the area under the $w(f)$ curve. This fact is somewhat surprising. It says that even though $I(t)$ may have the different appearances shown in Figs. 2–5, it is always normally distributed. For this reason it is often called "Gaussian noise," the terms "Gaussian distribution" and "normal distribution" being synonymous. This does not mean that all noises which have a power spectrum are Gaussian. An example to the contrary is a telegraph signal composed of marks and spaces chosen at random, the marks having a value of +1 and the spaces −1. This signal has a well-defined power spectrum but its values are limited to +1 and −1.

A distribution which is often confused with that of $I(t)$ is that of its envelope $R(t)$. As shown in Fig. 2, the envelope is of significance in problems involving noise in a narrow frequency band. $R(t)$ is also of importance because it represents the audio-frequency output of a radio detector when the input is radio-frequency noise. Choose a time t_1 at random and observe the value $R(t_1)$ of the envelope at that time. The probability that it will lie between the values R, $R+dR$ (chosen beforehand) may be shown to be

$$\text{Probability } [R<R(t_1)<R+dR] = \frac{RdR}{\sigma^2} \exp\left[-\frac{R^2}{2\sigma^2}\right]$$

where σ is, as before, the rms value of $I(t)$. R is never negative. This distribution is sometimes called the "Rayleigh distribution." Its integrated form is especially simple:

$$\text{Probability } [R<R(t_1)] = \exp\left[-\frac{R^2}{2\sigma^2}\right]$$

It will be observed, that just as in the case of $I(t)$, the distribution of $R(t)$ does not depend upon the shape of the power spectrum; it depends only on the area under $w(f)$. Fig. 6 shows the two probability distributions, the normal one for $I(t)$ and the Rayleigh for its envelope $R(t)$. The horizontal axis is marked off in units of σ, the standard deviation of $I(t)$.

A Flat-Bed Facsimile Telegraph Transmitter

W. D. BUCKINGHAM
NONMEMBER AIEE

AN experimental flat-bed facsimile telegraph transmitter with a number of novel features has been developed at the Water Mill Electronics Laboratory of The Western Union Telegraph Company. In the new machine, the message is placed on the flat bed of the transmitter and scanned by a series of closely spaced transverse lines as the sheet is advanced over the scanning line.

Such a device has several advantages over the more common type of facsimile transmitter which requires the message to be wrapped around a drum. With the flat-page or flat-bed transmitter, the message sheet requires no preparation or manipulation by the operator. It is simply laid on a table or dropped in a slot from which it is drawn by rollers through the machine. Furthermore, the sent message may be of any length. For example, copy from a telegraph teleprinter can be fed directly and continuously into the transmitter.

Optical System

The arrangement of the basic parts of a new machine is shown in the diagram of Fig. 1. Scanning is accomplished by sweeping a spot of light, 0.01 inch in diameter, horizontally across the message sheet from left to right at a uniform speed, and returning the spot quickly, at nineteen times the forward speed, to the left margin to repeat the cycle again and again as the message is advanced slowly at a uniform speed transversely across the scanning line. Light reflected from the message is collected by a cylindrical reflector, concentrated on a photocell, and converted into electric currents whose intensities are proportional to the reflected light at each instant.

TUNGSTEN ARC LAMP

The scanning light source is a miniature tungsten arc lamp, Fig. 2, developed specifically for facsimile use. It has a tungsten ball cathode, 10/1,000 inch in diameter supported on the end of a 4/1,000-inch-diameter tungsten wire, and a molybdenum disk anode. These two elements are mounted in a small glass bulb filled with argon gas.

The lamp operates from a special rectifier power unit which first delivers a high-voltage pulse to start the arc and then supplies a direct current of about 100 milliamperes at 40 volts to maintain the arc.

During operation the tungsten ball is maintained at incandescence by the heat of the arc to produce an extremely small, uniformly bright, stable, and well-defined point source of light. It can be focused with a single lens or mirror directly on the message in the facsimile transmitter without the use of the condensing lenses and pin-point aperture such as are required to produce an effective point source from the ordinary tungsten filament lamp usually employed in a "flying spot" scanner. The elimination of these elements by the use of the new lamp avoids

---

Paper 56-108, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 19, 1955; made available for printing November 29, 1955.

W. D. BUCKINGHAM is with the Western Union Telegraph Company, Water Mill, N. Y.

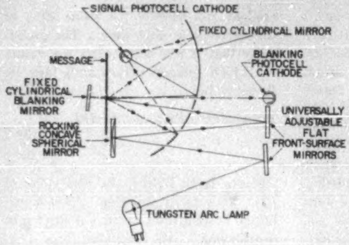

SIGNAL PHOTOCELL CATHODE
FIXED CYLINDRICAL MIRROR
MESSAGE
BLANKING PHOTOCELL CATHODE
FIXED CYLINDRICAL BLANKING MIRROR
UNIVERSALLY ADJUSTABLE FLAT FRONT-SURFACE MIRRORS
ROCKING CONCAVE SPHERICAL MIRROR
TUNGSTEN ARC LAMP

Fig. 1. Diagram of optical system of flat-bed facsimile scanner

expensive optics, results in a brighter scanning spot and reduces the length of the optical path.

The new tungsten arc lamp offers an additional advantage over the tungsten filament lamp for facsimile transmitter use in that the light output of the tungsten arc is relatively constant with changes in lamp current. This comparison is shown by Fig. 3. A 10-per-cent (%) increase in current causes the light output of the tungsten filament lamp to increase 93%. Under a similar current increase, the light output of the tungsten arc lamp increases only 7%. Thus, it is not necessary to operate the tungsten arc lamp from a voltage regulator as is required for the tungsten filament lamp in this service.

During the life of the tungsten arc lamp, the tungsten ball gradually evaporates until it is reduced to the diameter of the supporting wire. This occurs after 200 to 300 hours of use. The lamp can then be renewed by reversing the polarity and sending ten times normal current through it for an instant so as to form a new tungsten ball on the end of the wire. This cycle can be repeated several times before the lamp is discarded.

ROCKING MIRROR

In the new flat-bed transmitter, light from the tungsten arc is collected by a concave spherical mirror which then reflects the image of the incandescent tungsten ball in the lamp as a small bright spot of light on the face of the message. This mirror is a simple plano-concave lens, $1^3/_8$ inches in diameter, whose concave spherical surface has been coated with aluminum to form a concave mirror with a 20-inch radius of curvature. The mirror is placed so as to be substantially equidistant from the light and its image. Thus the optical system works at unity magnification and the high focal ratio of $14^1/_2$. Under these conditions it is possible to produce a very high-grade image of the source with very low-cost optics. Flat mirrors are interposed in the optical path between the lamp and the concave mirror, and again between the concave mirror and the message, to direct the light beam back and forth so as to reduce the over-all space required for the optical beam.

The concave mirror is mounted on supports so that it can be rocked through an 11-degree arc by a motor-driven cam. This motion causes the image to sweep across the message at a uniform speed during 95% of the time of one revolution of the cam and to return to the starting point in the remaining 5% of the time. The cam is made of nylon impregnated with molybdenum disulfide. This material combines the advantages of quiet operation and long life of the nylon with the low friction properties of the molybdenum compound. The mirror mount is made of aluminum that it may have little mass. As a result, it has been found possible to operate such a rocking mirror at speeds up to at least 30 times a second or 1,800 rpm. The cam-operated mirror with the associated motor, light chopper, speed-reducing gears, and clutch of a unit which operates at 180 rpm, is shown in Fig. 4.

CYLINDRICAL MIRROR AND PHOTOCELL

While the spot of light scans the message from left to right for a distance of about 8 inches and returns quickly to repeat the cycle, the message is advanced transversely across the scanning line at a speed which produces 100 scanning lines to the inch of the message sheet. As the spot scans across the sheet, the light reflected from the message is collected by a cylindrical reflector and brought to a focus as a line of light on the photocell cathode.

The photocell is of a type which has a long, narrow, light-sensitive surface so as to pick up as much as possible of the line of light generated by the cylindrical reflector. With this combination of cylindrical reflector and photocell, the whole length of the photo-cathode collects light all of the time regardless of the position of the spot of light along the scanning line and thus avoids the usual requirement for such service that the photo-cathode be equally sensitive at all points.

The cylindrical mirror collects the reflected light over an average solid angle of more than 90 degrees. If the reflected light is assumed to have a Lambert cosine spatial distribution, it can be shown that the cylindrical reflector collects more than 50% of the total reflected light. This collecting system is so efficient that an ordinary gas photocell provides ample signal output.

If the message being sent through the new machine is narrower than the scan line length, the scanning spot falls on a cylindrical mirror back of the message during the part of its sweep when it is off the message. This mirror sends the light into a photocell which operates to blank the inverter signal. This arrangement automatically provides a white background over the entire width of the received copy regardless of the width of the transmitted message.

Electrical System

A chopper wheel interrupts the light from the lamp at 2,700 cycles per second to produce a modulated signal from the photocell that can be handled by an a-c amplifier. This 2,700-cycle carrier frequency is sufficiently high to transmit the 1,200-cycle frequency produced by the 180-rpm scanner. A 5,400-cycle chopper is used on 360-rpm machines.

A resistance-capacity coupled amplifier is used to amplify the modulated signal from the photocell. The amplifier is equipped with a time-delayed automatic volume control which holds the output of the amplifier constant within 1/10 decibel as the input changes slowly over a 10-decibel range. This arrangement assures that the amplifier, over its range of control, will produce a constant signal from the message background, regardless of its color. Thus the amplifier adjusts itself automatically to the correct gain to produce a copy from the recorder which will have a white background and yet not fail to record light copy such as pencil marks that may be on the original message. The photocell, cylindrical reflector, and amplifier assembly are shown in Fig. 5.

Mechanical System

The various elements of the flat-bed scanner may be selected and arranged

Vol.
Pt. 1
1956

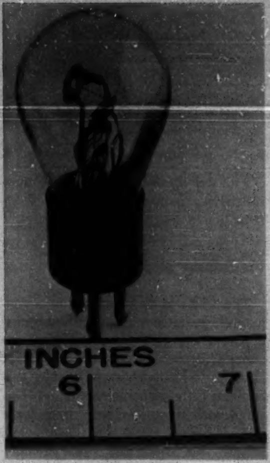


Fig. 2. Tungsten arc lamp developed for facsimile use

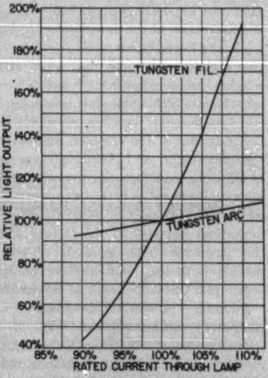


Fig. 3. Light versus current curve of tungsten filament lamp and tungsten arc lamp

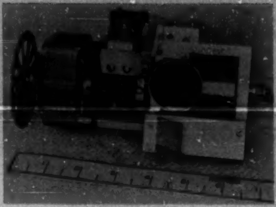


Fig. 4. Swinging mirror, cam, and light chopper assembly

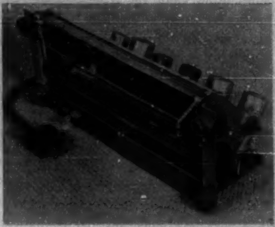


Fig. 5. Photocell, reflector and amplifier assembly

Fig. 6. Private-wire transmitter

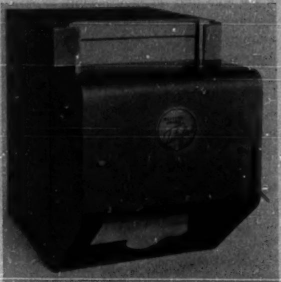


Fig. 7. Concentrator type of transmitter

in any convenient form to fit the requirements of the particular service in which it is to be used. Fig. 6 shows a machine designed to handle copy up to $8^1/_2$ inches in width for private-wire service. In use, the message is placed on the flat table of the machine with the feed roller resting directly above the part of the message at which transmission is to start. A pointer at the edge of the table is adjusted to the line on the message where transmission is to end. If a message longer than the standard 11-inch sheet is to be sent, the end-of-message pointer is drawn to the extreme end of its travel. The machine will then transmit the copy, whatever its length, until the end of the sheet has passed the scanning line.

When the machine starts, the sheet is fed at high speed to bring the message quickly to the scanning line. The rate of feed is then reduced to normal during transmission and at the end is again increased to feed the message sheet into the "sent" message compartment.

A flat-bed machine designed for Western Union concentrator service is shown in Fig. 7. Here, standard length 8- or $8^1/_2$-inch-wide telegrams are dropped into a slot on the front of the transmitter. If the message does not fill the entire message box, an indicator bar set just below the signature will cause the machine to terminate the transmission at that point. When the start signal is received, the message is time-stamped on its face automatically and then fed at high speed to bring the top of the message box to the scanning line. At the end of the transmission, the sheet is fed at high speed into the sent message box at the bottom of the machine.

Since the new experimental flat-bed transmitters require little of the operator, it becomes possible to equip the new transmitters with automatic message loaders. The operator then puts the messages to be sent into the automatic loader and it in turn delivers the messages to the transmitter as required, one at a time and in sequence. Such a device is under development.

Although models have been made of the transmitter described here, further experimentation and considerable laboratory and field testing will be done before any decision is made to undertake production of this type of Telefax device. Evaluation studies based on such tests are now in progress.

Application of a Digital Computer to the Design of Power Transformers to Specification

C. L. MOORE
ASSOCIATE MEMBER AIEE

W. T. DUBOC
ASSOCIATE MEMBER AIEE

P. A. ZAPHYR
NONMEMBER AIEE

MOST power transformers of 1,000 kva and larger are built to specification, and the problem of reducing the amount of time required to produce designs for them has been acute for many years. In early 1954, this problem became the focus of attention because of increasingly critical performance requirements which actually lengthened the time required for design while at the same time demands for shorter building cycles were becoming more and more pressing. Furthermore, developments in core steel and insulation increased the need for comprehensive studies of the effects of these improved materials on the economics of transformer design and construction. Such studies can be made on a theoretical basis. They can also be made by changing individual parameters (e.g., cross-sectional area of the core) and calculating the effect of this change on the transformer characteristics. However, neither of these methods is a wholly satisfactory substitute for making a series of designs all of which meet the same requirements, with one exception (e.g., noise level) which is varied over a specified range. Because a given charcteristic may have a greater effect on one rating than on another, the number of designs which must be made necessarily covers a wide range of ratings and is quite large. Not only is such a project expensive, tedious, and time-consuming, but the problem of consistency is acute since no two engineers will use exactly the same margins; in fact, the amount of margin which a single engineer thinks necessary may vary over a period of time.

One solution to the problem of reduction in design time seemed to lie in the possibility of programming a digital computer to produce designs with no more input than the specifications normally given a design engineer. The growing availability of such machines made an investigation of this possibility feasible.

At the outset, there was little experience on which to base the approach to such a project. There seemed to be two avenues available, both involving certain sizable disadvantages. The first, adaptation of the trial-and-error method used by the design engineer, in effect, would involve programming the engineer's judgment and experience into the computer. The second and much more ambitious procedure was a completely fresh approach to the whole design problem. In this case, one function such as cost would be selected as critical and, through use of techniques similar to those of linear programming, the design parameters (e.g., copper size and core dimensions) would be adjusted to minimize this function consistent with meeting the guaranteed specifications. However, any function such as cost which might be selected is an extremely complicated and, usually, nonlinear function involving all parameters, many of which are discontinuous. Copper strap, for example, is available only in discrete sizes. Furthermore, the necessary nonlinear programming techniques were not fully developed.

After consideration of these factors, it was decided that the more desirable approach was the adaptation of the design engineer's technique. Primarily responsible for this decision was the consideration that the problem was of sufficient urgency as to make a solution desirable in the foreseeable future. Furthermore, one of the main reasons for the project was the determination of the feasibility of using a digital computer in the commercial design of custom-built apparatus. As the project progressed, these two criteria, feasibility and the time element, were largely the basis for several decisions made. A second example is the limitation of the ratings to which the program would be applicable by the following restrictions:

1. Three-phase.
2. One thousand to 15,000 kva, inclusive.
3. High-voltage winding basic impulse level (BIL): 110 kv to 350 kv, inclusive.
4. Low-voltage winding BIL: 95 kv maximum.
5. Two windings.
6. No-load taps on high-voltage winding.

The activity of units covered by these restrictions is high and a method for reduction in design time was desired.

Preliminary Steps

The trial-and-error method of design involves an initial selection or assumption of parameters (e.g., copper strap and core dimensions) necessary to the calculation of the performance characteristics. Once calculated, these characteristics are compared with those specified. In general, this comparison will reveal a discrepancy between the two. The initial assumptions are then revised and the process is repeated as many times as necessary until an acceptable design is produced. In most cases, the specifications such as loss guarantee impose only upper limits on the characteristics of the transformer design while the lower limits result solely from the necessity of minimizing cost and weights; an exception is impedance which must be met within plus and minus limits. The final design is therefore one which, in the judgment of the engineer, minimizes cost and weight, yet has sufficient margin to allow for manufacturing variation and inherent inaccuracies in the basic data and which will also meet all restrictions and limitations imposed by the factory and the standards. The problem of consistency of designs mentioned previously arises from this large factor of judgment.

Adapting the method described to a computer is not a simple task. The first step is to make a mathematical outline complete with all necessary equations and logical decisions. The latter includes the comparisons of calculated results to guaranteed specifications and the steps to take in case of agreement and disagreement. In the following paragraphs points of particular importance are summarized.

Use of an expensive computer could be justified only, if the computer program were complete in itself with no intermediate logical decisions or interpretation required by an engineer or operator. For this reason the input to the machine, in addition to the program and the data concerning material characteristics, limitations, and dimensions which it incorporates, is as follows:

Paper 56-79, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 24, 1955; made available for printing December 12, 1955.

C. L. MOORE is with the Westinghouse Electric Corporation, Sharon, Pa.; W. T. DUBOC is with the Westinghouse Electric Corporation, Boston, Mass., and P. A. ZAPHYR is with the Westinghouse Electric Corporation, East Pittsburgh, Pa.

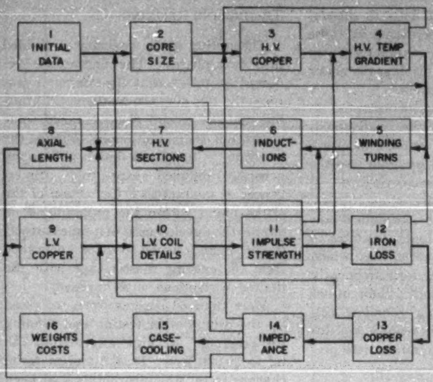


Fig. 1. Flow chart of computer program

1. Self-cooled and forced-air kva ratings.
2. High-voltage and low-voltage winding basic impulse levels.
3. Number and specified voltages of high-voltage winding taps and specified voltage of low-voltage winding.
4. High-voltage and low-voltage winding connections (delta or wye).
5. Noise level.
6. Temperature rise.
7. No load loss.
8. Exciting currents at 100% and 110% rated voltage.
9. Total loss.
10. Impedance.
11. Insulation space factor.
12. Eddy loss factors for high-voltage and low-voltage windings.

Items 1 to 10, inclusive, together with the restrictions just mentioned, define the operating characteristics of a transformer and are essentially the same as those received by the designer when he is given the job of designing a transformer. Item 11 was included because it is subject to possible slight variation from time to time, and item 12 was included because the method of calculating these factors was undergoing major revision at the time the program was developed.

It is evident that the initial values assumed for the various design parameters will have a considerable bearing on the number of repetitive trials (i.e., iterations) required to converge to an acceptable design. A design engineer usually has a wealth of previous designs from which he can select one with characteristics quite close to those of his current problem. It was considered impractical and expensive to store this tremendous number of data within the computer, so a substitute was developed in the form of a series of empirical equations based on a survey of recent designs covering the range of ratings to which the program was applicable. The most important was an equation for "Gross N" (defined as the product of the cross-sectional area of the core and the area of the core opening) as a function of the kva and BIL of the transformer.

The form in which basic data are most readily used by a designer is not the same as that which is most economical in terms of computer storage, and a large amount of adaptation was necessary. In general this involved development of equations to fit curves and tables wherever possible. In the case of data in discrete steps (e.g., sizes of copper strap), the tables were reduced to the minimum necessary.

In performing the iterative procedure of design, an engineer's judgment permits him to make partial calculations and adjustments in several parameters at once. For instance, he may know from experience that in a certain range of ratings, the no-load loss may govern induction whereas in another range, the noise level may be the determining factor. In another case he may know intuitively that a particular adjustment of a parameter which might be correct for, say high impedance cannot be used because it will result in high loss. The logic of a machine is such that it is restricted to a series of individual "either-or" decisions and it has no intuition at all, so the outline was prepared to allow for all possible contingencies. Also each time an adjustment in a parameter is made, all the characteristics which this parameter affects must be recalculated and retested against guaranteed values. This was perhaps the major problem. The final solution is best described by reference to Fig. 1, which is a block diagram of the flow of calculations through the program. From this it may be seen that, following establishment of design parameters in blocks 1 through 10, the first major characteristic which is calculated and tested is the impulse strength of the windings, block 11. If the design will meet the required impulse levels, the program proceeds to the calculation of iron loss. If it will not, corrections are made in one or more parameters as necessary, and previous calculations which are affected are recalculated. This process is repeated as many times as necessary until a design is obtained which will meet the required BIL. It should be noted that no consideration is given to other characteristics until the BIL requirement is satisfied, and this philosophy prevails through the program. The order in which the various characteristics are calculated was selected so that those least likely to fall outside the guarantees, under conditions defined by the initial assumptions and as determined from experience, are calculated first. The problem of oscillation described in the following was found to be reduced through use of this approach.

It is apparent from Fig. 1 that a given set of specifications may require a large number of iterations while the program converges to an acceptable design. During this process the problem of oscillation due to over-compensation may arise. For example, a given cross-sectional area of core may yield a design which meets all requirements except impedance, which is too low. To compensate, the area of the core may be decreased by one standard step but this, is turn, may result in an impedance which is too high. If this is compensated by increasing the area of the core by one step, oscillation may occur, since the original conditions will result. To prevent this from happening, the program is designed to test for return to the previous conditions, and should this happen that particular method of compensation is blocked and an alternate used.

To prevent overdesign (i.e., excessive margins) the program incorporates methods which will automatically attempt to design to maximum losses consistent with sufficient margins to assure that the resultant transformer will meet the required guarantees.

Once the mathematical outline and the supporting flow charts were developed, it was hand-tested. Recently designed typical transformers were selected and new designs were produced by working through the program by hand. This testing was of particular value in demonstrating inadequate consideration of possible contingencies as well as simple errors and omissions. The intermediate results obtained through this procedure were invaluable in later testing the computer program.

Computer Program

Having completed the preliminary steps of outlining, flow-charting and hand-testing the calculations, the way was prepared for programming a digital computer essentially to assume the role of a design engineer in processing a power transformer design.

CHOICE OF COMPUTER AND CODING SYSTEM

It was presupposed in preparing the mathematical outline that a stored program computer with large storage facilities would be employed. With the entire program in storage, the computer can choose among several alternate paths depending on the outcome of a prior calculation. Thereby, the flow required by the mathematical outline can be executed without external assistance. The IBM type-*701* computer was selected since it satisfied the requirements, in addition to being available on a service basis.[1]

Instead of coding the *701* in terms of its own language, it was decided to use a *701* interpretive routine, called Speedcode.[2] Speedcode permits one to perform all calculations without keeping track of the position of the decimal point, normally a cumbersome and time-consuming operation. It also provides the elementary functions (e.g., sine, logarithm), automatic checks, and other features designed to simplify the coding and testing of the program. These features, however, are achieved at the expense of increasing the calculation time. The decision to use Speedcode was based on the desire to establish the feasibility of the project in a reasonable length of time.

CODING

The *701* retains digital information in three different types of storage units: magnetic tapes, magnetic drums, and electrostatic tubes. Calculations can be performed only on numbers stored in electrostatic storage, so information on tapes and drums must be moved into electrostatic storage before being used.

It was apparent from the outset that the coding would be too large to be stored entirely in electrostatic storage and that it should be subdivided into smaller parts or subprograms. Toward that end, the program was separated into 16 subprograms, the divisions occurring at those points to which the flow could return in an iterative path. Fig. 1 displays the major subdivisions of the program.

Three features were incorporated in the organization of the coding: 1. Efficient utilization of storage facilities. 2. Fast-loading of the program into the computer by means of magnetic tape. 3. Simple running procedure.

As the coding of the mathematical outline progressed, it was found desirable to construct a set of subroutines which could be called on to perform frequently needed special functions. Subroutines for searching tables, for rounding calculated values to integer values, and for evaluating often-used equation were prepared.

To obtain the features of the program just mentioned the *701* storage is employed in the following manner:

1. In addition to being the working storage of the computer, electrostatic storage contains information required by almost all of the 16 subprograms. Included are 1. the input data listed in the section "Preliminary Steps"; 2. the rounding subroutines; 3. often needed constants; and 4. a set of electrostatic storage locations used to direct the flow, keep track of the number of iterations, and prevent nonconverging oscillations.

2. Magnetic drums contain the remaining subroutines, the tables, and the intermediate results.

3. One magnetic tape contains all 16 subprograms which are moved singly into the working area of electrostatic storage as called for by the flow. That tape also contains all information required initially by the drums and electrostatic storage, i.e. tables, subroutines, and constants. Therefore, it is the high-speed means of loading the program into the computer. Since the entire program is contained on that master tape, it must be preserved between runs.

4. Punched cards provide the primary source of information to the *701*. A set of cards is used in the program to load the design input data listed previously, and to transfer control to the master tape.

The master tape on receiving control first distributes the tables, subroutines, and constants to the drums and electrostatic storage as required. It then moves the first subprogram into the working area of electrostatic storage and causes control to transfer to that subprogram. From there the flow proceeds as directed by the mathematical outline. Since the master tape assumes the major role in starting the calculations, the operator need only place that tape in the proper unit and load the initial set of cards.

TESTING AND REFINING

The testing of the coding was divided into two phases. The first phase required assurance that the computer was properly following the mathematical outline and the second phase entailed some changes in design philosophy to refine the program. Changes in design philosophy generally entailed expediting convergences on guarantees. Correction of the errors found within the program necessitated the development of a tape correcting program designed to write a new tape from the existing master tape with the desired corrections superimposed.

More than 50 designs were processed by the computer in testing and refining the program. Ratings ranged from 750 up to 20,000 kva. Design and material changes were also incorporated to keep the program abreast of current design methods and to obtain experimental data.

DESCRIPTION OF FINISHED PROGRAM

The entire program of 3,600 Speedcode instructions and 1,000 data items is contained on a single magnetic tape. To process an order one simply punches the input data into six cards which, along with a deck of about 20 cards, are loaded into the *701*. From then on, the computer takes over completely, and in a matter of minutes prints out the same results previously obtained after days of effort by a design engineer. The computer also prints out enough information so that one can see exactly what calculations and iterations were performed in producing the design. In addition, a cost analysis is performed automatically at the end of each design.

The ease of running the program is evidenced by the fact that *701* computer operators were able to run designs completely with only one sheet of simple instructions. Designs have also been produced and revisions inserted on the master tape by using a transceiver link between the engineering offices and the computer installation. The transceiver is a device which can transmit and receive information in the form of punched cards by way of standard telephone circuits.

Before any computer design program can take an active role in the manufacturing process, it is imperative that some ready means be available to keep the program up to date with current developments in design procedure. For that reason the tape-correcting routine with its ease of operation forms an integral part of the transformer design program.

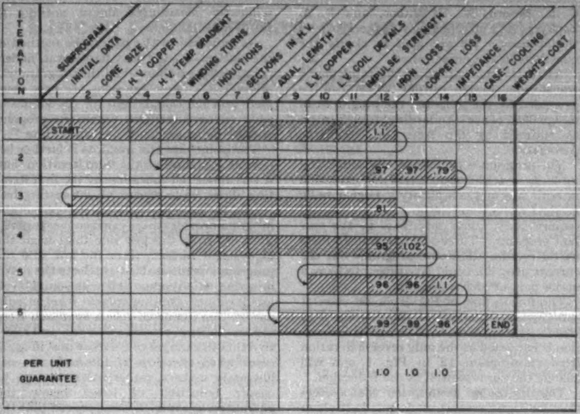


Fig. 2. Example of computer design

Example Showing Iterations

The design philosophy incorporated within the program can be demonstrated by following the iterations of a sample design. The iteration bars of Fig. 2 follow the flow as the computer iterates along the paths of Fig. 1 to design a 5,000-kva transformer with BIL of 350 kv. This example was chosen to demonstrate a variety of iterations taken to converge on design levels of losses and impedance. The first iteration found that the initial design exceeded the design level of iron loss in subprogram 12 by approximately 10%. To decrease the losses below the guaranteed values, the computer returned to subprogram 5 which lowered the induction by increasing the number of turns. On the second iteration both loss conditions were satisfied but the calculated impedance was only 79% of the guarantee. That low impedance necessitated a return to subprogram 2 where the core size was decreased, thereby increasing both turns and impedance on the following iterations. In this case the increase of turns resulted in iron losses below economically acceptable values, so a fourth iteration was taken to adjust the turns raising the iron losses nearer to the guarantee. However, high copper losses required a return to increase the copper size. The fifth iteration resulted in an impedance 10% over the guarantee. That slightly high impedance was corrected by a change of physical proportions during the sixth iteration and the design was completed. The computer time to accomplish this design including the cooling and cost calculations was 8.4 minutes. It should be noted that the computer has calculated many designs in less time with fewer iterations.

Analysis of Designs

The design program has successfully designed transformers within the range of 750 and 20,000 kva and with BIL including 350 kv. With an input of six punched cards the computer has determined the quantity, distribution, and design of materials totaling up to 94,000 pounds per unit. In general, the quality of the designs has proved excellent. The computer calculated designs have better balanced values of gradient and ampere-turn distribution than most comparable hand-calculated designs. Design time for the computer ranges from 3 to 12 minutes. This time could be materially reduced on some of the most recent digital computers.

Conclusions

Some of the results of significance to the engineering public and to the users and manufacturers of transformers are:

1. With only performance data as input, the computer automatically designs a transformer to meet all specifications in a 1-step process.
2. The 100-to-1 speed advantage of the computer decreases the engineering time requirements in processing transformer orders, thereby providing faster delivery.
3. A consistency of high-quality design is provided through the computer program which is based on the best engineering data available.
4. The use of computers in design will permit the manufacturer to pass benefits of new materials or new methods more rapidly to the user.
5. The development engineer has an opportunity to study the effect of varying any one parameter by obtaining from the computer a large number of consistent designs.
6. The use of computers will provide a fuller utilization of engineering manpower. By assuming routine design work, the computer can release design engineers from routine design to development work or further creative work.
7. The initial objective of providing the feasibility of automatic design was more than realized, and a valuable development and design tool has resulted.

References

1. The System Design of the IBM Type 701 Computer, W. Buchholz. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, 1953, pp. 1262–75.

2. The IBM 701 Speedcoding System, J. W. Backus. *Journal*, Association for Computing Machinery, New York, N. Y., vol. 1, 1954, pp. 4–6.

◆

Discussion

S. B. Williams and **P. A. Abetti** (General Electric Company, Pittsfield, Mass.): The authors are to be commended for their presentation. It would be interesting to know what percentage of the total power transformer design output can be processed with the use of the program described in the paper. Also, is this program suitable for extension to design units of larger size and complexity, and of higher basic insulation levels, e.g., 3-winding transformers and autotransformers?

To refer to Fig. 2, some steps are not very clear from the design standpoint. For instance, at the end of the third iteration it is found at routine 12 that the iron loss is 19% too low and the program switches to the induction routine (number 6). It is difficult to see how inductions can be changed while maintaining the same core size (routine 2) and the same winding turns (routine 5). In fact, change of the inductions will change the rated winding voltages which will no longer meet specifications. At the end of the fourth iteration, when it is found that copper loss is 2% high, an adjustment is made in the low-voltage winding only. This would seem to upset the balance between low-voltage and high-voltage winding rises. It is well known that optimum copper utilization may be obtained only when these rises are balanced.

Furthermore, it is not clear as to how a complete engineering design, as described by the authors in their oral discussion, can be obtained by establishing as fixed input data: 1. the insulation space factor, and 2. the eddy loss factor for the two windings. It would appear that these factors are not part of the customer's specifications or industry standards, which should be calculated from the actual dimensions of the core windings, rather than estimated in the initial stage. Unless the actual values coincide with the estimated values, the test results may not agree with the calculated results and with the customer's specifications.

It appears that the design in Fig. 2 does not converge smoothly and logically to the desired final answer. There are evident indications of overcompensation of certain characteristics which introduce abnormal oscillations. For instance, at the end of the first iteration the only characteristic which has been computed, the iron loss, is 10% off. After the third iteration the same characteristic, again the only one computed, is 19% off in the opposite direction. From this it would appear possible that under certain unfavorable circumstances, the design might oscillate back and forth about the desired value, without ever converging to that value. In the paper[1] presented by the discussers it was recognized that convergence must be determined to include all possible variables, and that appropriate damping factors must be established to avoid overcompensation. Fig. 2 of reference 1 shows clearly that oscillations are avoided in this manner. The question of convergence becomes more critical as the complexity of the design increases, the requirements become more exacting, and the number of design variables increase. Such is the case of 3-winding transformers and autotransformers for extra-high voltage with difficult tap arrangements in several windings which are included in the computer program we have described.[1]

REFERENCE

1. APPLICATION OF DIGITAL COMPUTERS TO TRANSFORMER DESIGN, S. B. Williams, P. A. Abetti, E. F. Magnusson. *AIEE Transactions*, vol. 75, pt. III, 1956 (*Paper No. 56-80*).

C. L. Moore, W. T. Duboc, and P. A. Zaphyr: The discussion is appreciated by the authors. We are glad to clarify the questions in the order raised.

As to the percentage of the total power transformer design output which can be processed by the program, the program can design 62% of the standardized transformers as outlined by the American Standards Association.

The program is fundamentally sound and provides a sound foundation for extension in both size and complexity. Higher basic insulation levels and autotransformers would require only minor additions to the current program. Three-winding transformers would require additional programs but the current program could be utilized to form a major part of such a program.

The fourth flow of Fig. 2 should start at subprogram 5 instead of subprogram 6 as erroneously indicated. This 12-to-5 iteration is indicated as the only reverse iteration from subprogram 12 in Fig. 1 and will change the turns as well as the induction.

The discussers' assumption that corrections for high losses are made only in the low-voltage coil is unjustified, though admittedly such an inference might be drawn from Figs. 1 and 2. However, reduction of copper loss with constant turns maintained can be achieved by an increase in the thickness of the conductor which only affects the radial dimensions of the conductors and coils. Thus, reduction of losses in either winding can be accomplished by changes of radial dimensions only within subprogram 13 so that return to subprogram 10 is required simply to correct diameters and to verify that these changes did not adversely change the impulse strength and iron losses. The program maintains a very good balance of winding rises by reducing the losses of the winding with the higher calculated rise. An indication of the balance attained and the optimum utilization of copper is demonstrated by the values of the calculated winding gradients. In the example cited the high-voltage temperature gradient calculated to be 5.7 degrees centigrade and the low-voltage gradient to be 5.6 degrees.

Our term "insulation space-factor" has proved misleading. The term does not refer to any over-all ratio of insulation to conductor area as interpreted by the discussers, but merely indicates the compression factor which is applicable to the nominal value of the axial build-up of radial spacers in the high-voltage column. The inclusion of this factor was only for development consideration. In production designs it is loaded as 0.94 for all sizes and ratings and consequently contributes almost nothing to the design philosophy.

As indicated in the paper, the method of calculation of eddy loss factor was undergoing major revision when the program was originated. These revisions are now complete and the calculations for the values of eddy currents included within the program and are no longer loaded as input data.

The point that the iron loss is farther below the guarantee on the third iteration than it was above the guarantee on the first iteration allows us to discuss an interesting feature of the program. The philosophy of the program provides more rapid convergance of iron loss from the low side than from the high side. In fact, from the low side the program in almost all cases selects the maximum induction compatible with sound level, loss level and exciting current level in only one 12-to-5 iteration. This approach from the low side has proved so certain and effective that in changes of core size and in some cases when iron loss is too high, we deliberately make a rather coarse change to ensure reduction of losses below the guarantee, as experience has proved that the program will seek the optimum induction on the next iteration. Thus, oscillation about the iron loss guarantees have been of negligible concern with this program. Regarding the general problem of oscillation, it may be mentioned that the only point in the program at which this problem has arisen is in connection with the procedures designed to correct for high and low impedance. As pointed out in the body of the paper, the program is designed so that a continuous check is maintained as to whether or not oscillation is occurring. Since there are several ways to adjust a design in order to bring the impedance within tolerable limits, oscillation simply causes the machine to switch from the preferred to the first alternative method of correction. All that is lost is the time required for two iterations which is insignificant especially since the problem arises only infrequently. It should be noted that every characteristic required of a transformer design is similar to impedance in that the design parameters can be adjusted in several ways in order to make the calculated value equal the guaranteed value. Hence, should the problem of oscillation arise anywhere in the program, it can always be solved by the method described here and in the paper.

REFERENCE

1. REQUIREMENTS FOR TRANSFORMERS. *ASA C57.12a-1954*, American Standards Association, New York, N. Y., 1954.

Linear Programming on an Electronic Analogue Computer

INSLEY B. PYNE
ASSOCIATE MEMBER AIEE

A GREAT deal of attention has been given in recent years to the computational aspects of linear programming. The reason for this development lies in the applications of the problem to such fields as operations research, game theory, logistics, military tactics, and others. The object of this paper is to present a method for performing linear programming on an analogue computer containing only standard components. The size of the problem that can be solved is limited by the number of computing amplifiers available and by the number of inputs to each amplifier. The method is unusually flexible because the computer constantly seeks new solutions as the parameters of the problem are varied.

An Optimization Problem

Although the method is capable of solving several types of optimization problems, it will be illustrated by a single type which can be stated in the following form. It is required to maximize the objective function of n variables

$$Z=\sum_{k=1}^{n}\frac{x_k}{\alpha_k} \qquad (x_k \geq 0) \tag{1}$$

subject to the m restrictions

$$\begin{aligned}&\sum_{k=1}^{n}\frac{x_k}{a_{1k}}\leq 1\\ &\sum_{k=1}^{n}\frac{x_k}{a_{2k}}\leq 1\\ &\sum_{k=1}^{n}\frac{x_k}{a_{mk}}\leq 1\end{aligned} \tag{2}$$

where the α's and a's are constants. The method described in this paper solves the problem with $(n+2m+2p)$ amplifiers, where

n = number of variables
m = number of restrictions
p = number of negative quantities among the a's

The reasons for using denominator coefficients will become apparent later.

Geometrical Interpretation

The variables $x_1, x_2, \ldots, x_n$ can be treated as the co-ordinates of a point in a Euclidean space of n dimensions. The point with these co-ordinates will be called the objective point. The objective function, equation 1, is continuous and single-valued everywhere within this space and therefore can be used to define a gradient (grad) vector, grad Z. If $\mathbf{i}_1, \mathbf{i}_2, \ldots \mathbf{i}_n$ are unit vectors in the directions of the co-ordinate axes, then

$$\text{grad } Z=\nabla Z=\sum_{k=1}^{n}\frac{\partial Z}{\partial x_k}\mathbf{i}_k=\sum_{k=1}^{n}\left(\frac{1}{\alpha_k}\right)\mathbf{i}_k \tag{3}$$

The vector ∇Z is everywhere constant, normal to the hyperplanes of equal Z, and in the direction of steepest ascent.

The jth restriction of equations 2 represents a hypervolume in the n-space bounded by the hyperplane

$$\sum_{k=1}^{n}\frac{x_k}{a_{jk}}=1$$

This bound is called the jth edge. An edge separates the region of the space in which an inequality is satisfied from the region in which it is violated. The space in which all of the given inequalities are satisfied is called the allowed region. The allowed region is ordinarily closed in the sense that the objective function cannot increase without limit without entering the restricted region. The scalar a_{jk} represents the intercept of the jth edge with the kth co-ordinate axis. Any of the inequalities, equation 2, are superfluous if its edge lies wholly within the restricted region.

Define a vector $\mathbf{N}_j$ normal to the jth edge and directed toward the allowed region. There are exactly m such vectors and each is associated with a particular restriction. To compute the components of $\mathbf{N}_j$, first note that the following linearly independent vectors are parallel to the jth edge. i.e., they are on the hyperplane

$$\sum_{k=1}^{n}\frac{x_k}{a_{jk}}=1$$

$$\begin{aligned}&(a_{j1}\mathbf{i}_1-a_{j2}\mathbf{i}_2)\\ &(a_{j2}\mathbf{i}_2-a_{j3}\mathbf{i}_3)\\ &(a_{jn-}\mathbf{i}_{n-1}-a_{jn}\mathbf{i}_n)\end{aligned} \tag{4}$$

The vector $\mathbf{N}_j$ has the properties that its inner product with each vector of the set, equation 4, vanishes and also that it is directed away from the jth edge toward the allowed region. Therefore it can be written as

$$\mathbf{N}_j=-\left(\frac{\mathbf{i}_1}{a_{j1}}+\frac{\mathbf{i}_2}{a_{j2}}+\ldots+\frac{\mathbf{i}_n}{a_{jn}}\right) \tag{5}$$

Next define m quantities $\delta_1, \ldots, \delta_m$, one for each inequality, such that

$$\begin{aligned}&\delta_j=0 \quad \text{when} \sum_{k=1}^{n}\frac{x_k}{a_{jk}}\leq 1\\ &\delta_j=1 \quad \text{when} \sum_{k=1}^{n}\frac{x_k}{a_{jk}}> 1\end{aligned} \tag{6}$$

Define the vector $\mathbf{f}$ as follows

$$\mathbf{f}=K \text{ grad } Z+\sum_{j=1}^{m}\delta_j\mathbf{N}_j \tag{7}$$

where K is a constant.

In terms of components, f may be written

$$\mathbf{f}=f_1\mathbf{i}_1+f_2\mathbf{i}_2+\ldots+f_n\mathbf{i}_n \tag{8}$$

From equations 3, 5, and 7 the kth component is

$$f_k=\frac{K}{\alpha_k}-\sum_{j=1}^{m}\frac{\delta_j}{a_{jk}} \tag{9}$$

Both the vector $\mathbf{f}$ and its kth component f_k depend on the co-ordinates of the objective point because of the presence of the δ's in equations 7 and 9.

Let the position of the objective point be described by the vector

$$\mathbf{r}=x_1\mathbf{i}_1+x_2\mathbf{i}_2+\ldots+x_n\mathbf{i}_n \tag{10}$$

The velocity of the objective point in the n-dimensional space may therefore be defined as the vector

---

Paper 56-147, recommended by the AIEE Committee on Computing Devices and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 4, 1955; made available for printing January 4, 1956.

INSLEY B. PYNE is with Princeton University, Princeton, N. J.

The assistance of the following people is gratefully acknowledged: L. P. Neuwirth, W. C. Ridgway, Prof. W. C. Johnson, and Prof. W. H. Surber.

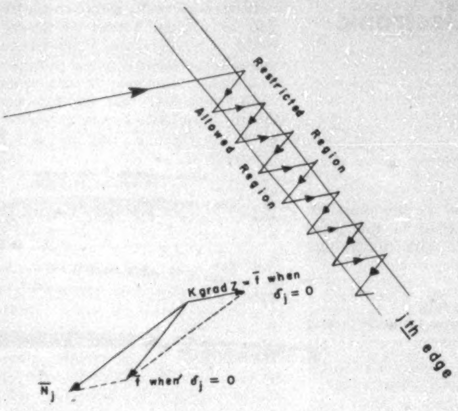


Fig. 1 (above). Motion of objective point in neighborhood of a restriction

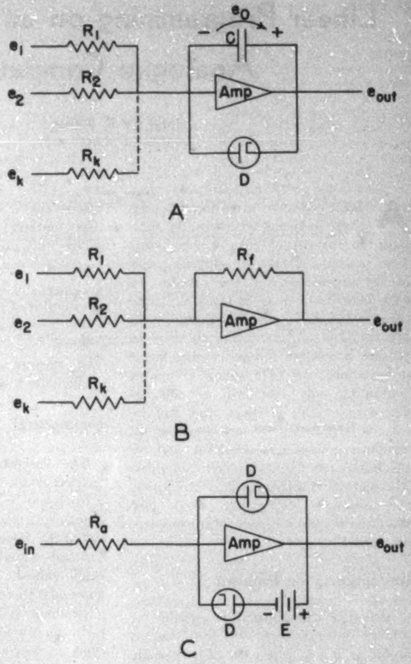


Fig. 2 (right). The three basic elements for linear programming

$$\mathbf{v}=\frac{d\mathbf{r}}{dt} \qquad (11)$$

where t is the time.

The equation which describes the method of this paper can now be written. Define a constant scalar γ such that

$$\frac{d\mathbf{r}}{dt}=\gamma\mathbf{f} \qquad (12)$$

Equation 12 describes the motion of the objective point as the solution is approached. The solution is to be the value which $\mathbf{r}$ finally attains. From equation 9 the kth component of equation 12 may be written

$$\frac{dx_k}{dt}=\gamma\left(\frac{K}{\alpha_k}-\sum_{j=1}^{m}\frac{\delta_j}{a_{jk}}\right) \qquad (13)$$

The objective point moves through the allowed region with a velocity γK grad Z, until it reaches a restriction such as the jth. At this point, the motion is determined by two vectors, the gradient and the vector $\mathbf{N}_j$ normal to the jth hyperplane. If N_j is greater than the normal component of K grad Z, the point will be ejected from the restricted region. The actual rebound is infinitesimal in magnitude as the objective point moves into the allowed region. The vector $\delta_j\mathbf{N}_j$ then vanishes, and the gradient causes the motion to reverse until the point again enters the restricted region. In this way, the objective point moves along the restriction boundary in the direction of the projection of the vector gradient Z on the jth hyperplane. This motion is shown greatly exaggerated in Fig. 1.

To prevent breakthrough into the restricted region, we must have

$$-K \text{ grad } Z\cdot\mathbf{N}_j<N_j^2 \qquad (14)$$

With the use of equations 3 and 5, this determines an upper limit on K

$$K<\frac{\displaystyle\sum_{k=1}^{n}\frac{1}{a_{jk}^2}}{\displaystyle\sum_{k=1}^{n}\frac{1}{a_{jk}\alpha_k}} \qquad (15)$$

which must be satisfied for every j.

Solution on Analogue Computer

To solve equation 13 on an analogue computer, it is necessary to divide the computer into the following three types of systems.

1. A set of n summing integrators limited to allow positive outputs only.
2. A set of m summing amplifiers, one for each restriction.
3. A set of m switches, one for each restriction.

The positive-output summing integrator is shown in Fig. 2(A) and, for positive e_{out} only, it satisfies the equation

$$e_{out}=-\frac{1}{C}\int\left(\frac{e_1}{R_1}+\frac{e_2}{R_2}+\ldots+\frac{e_k}{R_k}\right)dt+e_0 \text{ volts} \qquad (16)$$

where e_0 is the initial voltage across the condenser C. The diode D prevents the output from becoming negative.

The summing amplifier is shown in Fig. 2(B) and satisfies the equation

$$e_{out}=-R_f\left(\frac{e_1}{R_1}+\frac{e_2}{R_2}+\ldots+\frac{e_k}{R_k}\right) \text{ volts} \qquad (17)$$

The switch is shown in Fig. 2(C). Its properties are described as follows

$$e_{out}=0 \text{ volts} \quad \text{when } e_{in}>0 \qquad (18)$$

$$e_{out}=E \text{ volts} \quad \text{when } e_{in}<0$$

Fig. 3 shows the interconnection of n integrators, m summers, and m switches. The output of the kth integrator may be found by equation 16

$$e_k=-\frac{1}{C}\int\left(-\frac{\beta B_k}{R_k}+\frac{V_1}{r_{1k}}+\frac{V_2}{r_{2k}}+\ldots+\right.$$

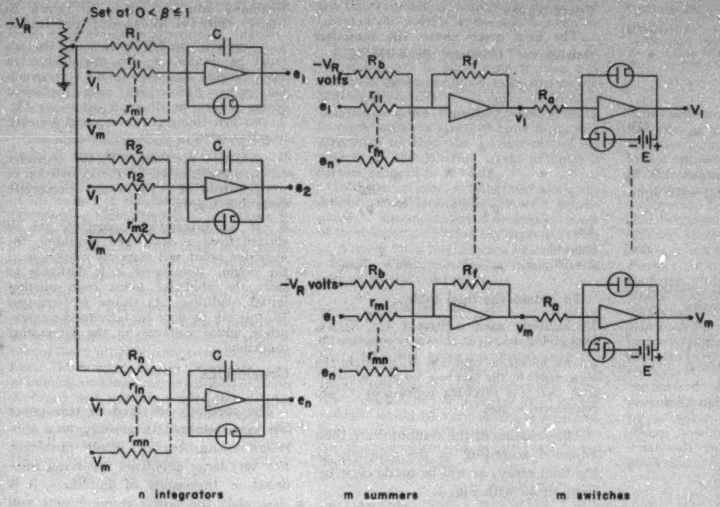


Fig. 3 (left). Connection of computer for linear programming

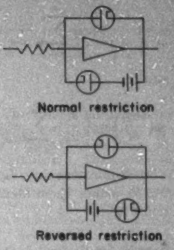


Fig. 4 (above). Alternate forms of switching device

$$\left.\frac{V_m}{r_{mk}}\right)dt \text{ volts} \qquad (19)$$

if the condenser C is initially discharged. The output of the jth summing amplifier is

$$v_j = -R_f\left(-\frac{V_R}{R_b}+\sum_{k=1}^{n}\frac{e_k}{r_{jk}}\right) \text{ volts} \qquad (20)$$

The jth switch has an output voltage V_j such that

$$V_j = 0 \text{ volts} \qquad \text{when } v_j > 0 \qquad (21)$$

$$V_j = E \text{ volts} \qquad \text{when } v_j < 0$$

Let δ_j be defined as follows

$$\delta_j = 0 \qquad \text{when } \sum_{K=1}^{n}\frac{e_k}{r_{jk}} \le \frac{V_R}{R_b} \qquad (22)$$

$$\delta_j = 1 \qquad \text{when } \sum_{K=1}^{n}\frac{e_k}{r_{jk}} > \frac{V_R}{R_b}$$

It follows that the output voltage of the jth switch is

$$V_j = \delta_j E \text{ volts} \qquad (23)$$

Differentiating equation 19 with respect to time, and introducing the symbolism just given, the equation which describes the computation performed by the circuit of Fig. 3 may be written

$$\frac{de_k}{dt} = \frac{1}{C}\left(\frac{\beta V_R}{R_k} - E\sum_{j=1}^{m}\frac{\delta_j}{r_{jk}}\right) \text{ volts per second} \qquad (24)$$

Equation 24, which describes the operation of the computer network of Fig. 3, is identical in form with equation 13, which describes the desired motion of the objective point of the function Z. Therefore, with proper selection of the computing resistors, the computer is properly connected to perform the computation of equation 13.

Completion of Analogue: Scaling

Each analogue computer operates best within the ranges of time and voltage which are implicit in its design. The following quantities, which determine these ranges, are associated with the computing equipment:

1. Integrating condenser C, microfarads.
2. Reference voltage V_R, volts.
3. Battery, E, volts.
4. Maximum allowable amplifier output voltage, $E_{\max}$, volts.
5. Range of computing resistances.

$R_{\min} \ldots R_{\text{mid}} \ldots R_{\max}$, megohms

A usual value for the integrating capacitors is 1 microfarad, except in high-speed repetitive computers, which use a value smaller than this.

The voltage E can be supplied by a dry battery and its value is not critical. The most convenient voltage is one that is of the same order of magnitude as the computer reference voltage. Since the battery draws negligible current, a small unit may be used.

The resistance R_{mid} can have a value anywhere within the upper middle part of the available range. Whatever value of resistance is recommended for use in the feedback and input paths of a sign-changing amplifier is a convenient choice. Alternately, R_{mid} may be chosen so as to make the product (CR_{mid}) either 1 second or some other convenient unit of time. On a computer using 1-microfarad integrating capacitors, a convenient value for R_{mid} is 1 megohm.

The resistance R_a is not critical, but it should be much larger than the forward resistance of the diodes. A convenient value is R_{mid}. The resistances R_f and R_j should be of the same order of magnitude as R_{mid}. Equation 20 shows that, if the e's are all initially zero, the output of each summing amplifier will start at $E_{\max}$ if R_f and R_b are chosen in the ratio

$$\frac{R_f}{R_b} = \frac{E_{\max}}{V_R} \qquad (25)$$

To establish the relation between equation 13 and its analogue, equation 24, define a set of scale factors S_k such that

$$e_k = S_k x_k \qquad k = 1, 2, \ldots, n \qquad (26)$$

Substituting for x_k in equations 6 and 13, these become

$$\delta_j = 0 \qquad \text{when } \sum_{k=1}^{n}\frac{e_k}{S_k a_{jk}} \le 1 \qquad (27A)$$

$$\delta_j = 1 \quad \text{when} \sum_{k=1}^{n} \frac{e_k}{S_k a_{jk}} > 1 \tag{27B}$$

and

$$\frac{1}{S_k}\left(\frac{de_k}{dt}\right) = \gamma\left(\frac{K}{\alpha_k} - \sum_{j=1}^{m} \frac{\delta_j}{a_{jk}}\right) \tag{28}$$

Multiplication of equations 27 by V_R/R_b makes them identical with equations 22 if

$$r_{jk} = \frac{R_1}{V_R} S_k a_{jk} \tag{29}$$

Substituting for a_{jk} from equation 29 into equation 28, we may write

$$\frac{de_k}{di} = \frac{\gamma R_b S_k^2}{E V_R}\left(\frac{KEV_R}{R_b S_k \alpha_k} - E \sum_{j=1}^{m} \frac{\delta_j}{r_{jk}}\right) \tag{30}$$

This is identical with equation 24 if

$$R_k = \mu S_k \alpha_k \tag{31}$$

where $\mu = \beta R_b/KR$

and also

$$\gamma = \frac{EV_R}{CR_b S_k^2} \tag{32}$$

Through the relations 29, 31, and 32 the analogy is now complete.

A convenient method for proceeding is as follows. Write an array, such as that for the unscaled problem below, which represents equations 1 and 2 of the original problem. In each column, select the largest coefficient, max a_{jk}, and determine the scale factor for that column from equation 29, using R_{mid} for the corresponding max r_{jk}

$$S_k = \frac{V_R R_{\text{mid}}}{R_b(\max a_{jk})} \tag{33}$$

Multiply each column in the table by itsscale factor, obtaining an intermediate form shown in the intermediate array.

ARRAY FOR UNSCALED PROBLEM

	x_1		x_n
N_1	a_{11}	…	a_{1n}
·	·		·
·	·		·
·	·		·
N_m	a_{m1}	…	a_{mn}
Z	α_1	…	α_n

INTERMEDIATE ARRAY

	$e_1 = S_1 x_1$	…	$e_n = S_n x_n$
	$S_1 a_{11}$	…	$S_n a_{1n}$
	·		·
	·		·
	·		·
	$S_1 a_{m1}$	…	$S_n a_{mn}$
	$S_1 \alpha_1$	…	$S_n \alpha_n$

FINAL ARRAY

The final array shows the computer resistances. Compare this with Fig. 3.

	e_1	…	e_n
V_1	$r_{11} = \frac{R_b}{V_R} S_1 a_{11}$	…	$r_{1n} = \frac{R_b}{V_R} S_n a_{1n}$
·	·		·
·	·		·
·	·		·
V_m	$r_{m1} = \frac{R_b}{V_R} S_1 a_{m1}$	…	$r_{mn} = \frac{R_b}{V_R} S_n a_{mn}$
Analogue of $Z = Z^*$	$R_1 = \mu S_1 \alpha_1$	…	$R_n = \mu S_n \alpha_n$

To obtain the final table:

1. Multiply each a-element by R_b/R_R, thus giving the r_{jk}, as shown by equation 29.

2. As shown in equation 31, multiply the α-elements of the last row by any constant μ so that the resulting resistances R_k are convenient in size.

The resistors of the computer are then adjusted according to the schedule of the final array, as will be made clear by comparison with Fig. 3.

In the expression $\mu = \beta R_b/KE$, the quantities R_b and E have already been selected. The factor K determines the magnitude of the influence of grad Z, as shown by equation 7, and hence influences the speed with which the machine approaches the solution. Once μ is fixed by the use of convenient values for the R_k, the potentiometer setting β, Fig. 3, determines the value of K. If K is made larger than the limit indicated by the relation 15, the objective point will break through a barrier. This can be detected by the overloading of an integrator, and can be corrected by reducing β.

The velocity of motion of the objective point is directly proportional to the factor γ of equation 12, which is given by equation 32.

Additional Aids to Computation

Various additional aids in computation are as follows:

1. To move an edge parallel to itself inward, lower the voltage which is applied to R_b of the corresponding summer from a value V_R to some smaller voltage; and to move it outward, raise the voltage applied to the corresponding summer and lower the corresponding resistance R_f proportionately to prevent saturation of the summing amplifier; see equation 25.

2. To minimize instead of maximize an objective function: apply $+V_R$ instead of $-V_R$ to the integrators.

3. To interchange the restricted and allowed regions for any restriction edge: reverse input and output terminals of the switching amplifier. This is shown in Fig. 4. Only two wires need to be changed.

4. Negative coefficients among the a's must be treated by the introduction of sign-changing amplifiers in the corresponding paths. Since each such coefficient enters the wiring diagram twice, see Fig. 3, two sign changers are required for each negative term.

5. Negative coefficients in the objective function are handled by the application of $+V_R$ instead of $-V_R$ at the appropriate integrator inputs.

6. If the integrating capacitors are all started from a discharged condition, the objective point will start its motion from the origin. Sometimes it is desirable to start the objective point from another initial condition. As shown by equation 16, this can be done by introducing appropriate initial voltages on the integrating capacitors.

Conclusions

The method described in this paper has been applied successfully to a considerable number of small problems. For very large problems involving hundreds or thousands of amplifiers, it is clear that the power requirements will become excessive. For such problems the development of solid-state amplifiers offers considerable promise. A special problem board would be convenient. Since each resistance among the a's enters the problem twice, these could be combined in a linear programming computer so that both are set in a single operation. The entire computer could be internally wired, and its programming board would then be a set of knobs arranged and marked as in the final array given previously.

However, in its present form, problems of moderate complexity can be solved readily on commercially available computers. Even problems involving certain types of nonlinearities may be introduced without drastic alteration of the method. The flexibility inherent in the method will perhaps be of benefit in guiding the intuition of the theorist in the field of optimization.

◆

Discussion

Philip Wolfe (Princeton University, Princeton, N. J.): The computational method Professor Pyne has developed promises to increase greatly the utility of the linear programming formulation of certain classes of operational problems.

Its use seems to be supplementary to that of the simplex method of George Dantzig, designed for use on digital computing machinery for the maximization of a linear function of variables subject to linear inequalities.[1] Dantzig's method gives, to an

Vol. 7
Pt. 1
1956

extreme degree of accuracy, answers which have other quite useful properties; but the large number of computations involved (about $5m^3$ multiplications) entails round-off problems, and makes the solving of a reasonably sized program ($m=50$) a matter of several hours. Moreover, a change in the given data usually requires a rerun of the whole problem.

The analogue method will be especially useful in removing this last difficulty. In many practical problems the data are known with so little accuracy (much less than that of the values of the electrical components representing them) that it is important to be able to explore the solutions of a whole class of problems with data near those of the given problem. Since the author's method allows these changes to be made and solutions to be found quickly and examined for "reasonableness," uncertainty regarding data should no longer present as great an obstacle as heretofore in the application of linear programming to realistic problems. Moreover, certain nonlinear programming problems can be phrased as linear problems with variable coefficients, so that analogue methods may be useful here, too.

It seems difficult to formulate a precise general statement about the length of time required to obtain a solution by this method, but certain practical bounds on the rate of travel of the point x to the solution can be indicated.

Denoting by K the speed of x when no constraints are in play, suppose that x is constrained to move at an angle θ to its unconstrained direction. It is easy to see that the effect of the constraints will cause x to move with the speed $K \cos \theta$ in its new direction, so that the component of its velocity in the direction of the gradient is further reduced to $K \cos^2 \theta$.

In the worst case, x will travel along the 1-dimensional intersection of $n-1$ hyperplanes. We can obtain an average estimate of its velocity by finding the expected value of $K \cos^2 \theta$ for a line selected at random in n-space. The calculation is straightforward and yields the expected value

$$\overline{K \cos^2 \theta}=\frac{\int K \cos^2 \theta dS}{\int dS}=\frac{K}{n}$$

the integration being performed over the surface of the unit sphere in n-space. Although this number is small when n is large, few problems may be expected in which the solution would be very slowly approached, because this case will occur only toward the end of the problem and, in any event K can be made quite large.

Reference

1. Activity Analysis of Production and Allocation (book), edited by Tjalling C. Koopmans. John Wiley & Sons, Inc., New York, N. Y., 1951, pp. 339–47.

Insley B. Pyne: Dr. Wolfe's calculation for the order of magnitude of the speed of the objective point in n-space is interesting and, in my opinion, optimistic. The optimism arises from the assumption that the speed of the point along a single constraint is proportional to the cosine of the angle between the directions of constrained and unconstrained motion. Actually, the motion involves a saw-toothed path back and forth through the constraint so that the resulting component of the velocity along the constraint is smaller than Dr. Wolfe's estimate. As the vector field normal to the hyperplane is increased in magnitude, the return velocity increases and the constrained speed approaches Dr. Wolfe's value as a limit. Circuitwise, this corresponds to increasing the battery voltage E and/or multiplying the resistances r_{ji} to the inputs of the integrators by a factor less than one. Experiment shows, however, that there are limits to what can be done accurately by these methods.

Properties of a D-C Arc in a Magnetic Field

L. P. WINSOR
MEMBER AIEE

T. H. LEE
MEMBER AIEE

MAGNETIC fields are used extensively in practice to help control the position or motion of an electric arc. Properly applied, they can be used as an aid to either the stabilization or extinction of an arc. Improperly controlled, they may introduce unwanted instability to an arc.

In spite of the wide application of magnetic fields to electric arcs, however, only meager knowledge of the fundamental behavior and mechanism of the motion of an arc in a magnetic field is available. Stolt[1] studied the rotation of an arc in a magnetic field but his work was limited to low-current arcs under 10 amperes. More recently, a number of investigations have been made of the motion of a low-pressure arc in a magnetic field in an effort to explain the retrograde motion of the arc.[2] Conclusions based on these results cannot safely be extrapolated to apply to high-pressure high-current arcs encountered in switchgear and arc welding.

This paper describes the behavior of d-c arcs at atmospheric pressure with currents up to 100 amperes when subjected to uniform magnetic fields up to 200 gauss. Electrodes studied include silver, copper, aluminum, carbon, tungsten, molybdenum, nickel, and titanium.

Experimental Procedure

The magnetic field was produced by a solenoid 3 feet in length by 12 inches in diameter. The solenoid was wound with 2,337 turns of wire in four layers on a form of glass-fiber cloth bound by Permafil.* End-turn correction was provided so that the field midway along the length was practically uniform throughout the cross section. The electrodes and contact-making mechanism were mounted on a textolite disk, 12 inches in diameter, which fitted snugly inside the solenoid.

The details of the electrode assembly are shown in Fig. 1. The electrodes were square rods 1/4 inch by 1/4 inch mounted parallel to each other along the diameter of the disk. The gap which could be adjusted by means of mounting screws was normally set at 1/8 inch. For copper and carbon electrodes, 1/4-inch commercial bars were used. Electrodes of other materials were brazed directly to the copper bars. The titanium strips were attached with the help of the General Electric Research Laboratory through the courtesy of Dr. Cobine. With the exception of carbon, all electrodes were polished either mechanically (copper, silver, nickel, aluminum) or electrically (tungsten and molybdenum). In general, the electrodes were mounted with their flat surfaces parallel. The edge mounting shown in Fig. 1 was an exception used during some of the preliminary tests.

---

* General Electric trade-name for a polyester resin.

Paper 56-132, recommended by the AIEE Committee on Basic Sciences and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted May 16, 1955; made available for printing November 21, 1955.

L. P. Winsor is with Rensselaer Polytechnic Institute, Department of Electrical Engineering, Troy, N. Y., and T. H. Lee is with the General Electric Company, Schenectady, N. Y.

This paper is based on the experimental part of a doctor's thesis presented by Mr. Lee to Rensselaer Polytechnic Institute, May 1954.

The authors wish to express their appreciation to J. W. Cooke, H. B. Rhoads, and T. K. Knapp, General Electric Company, for their support of this investigation; to J. H. Steadwell for his assistance in designing the test facilities; to J. H. Connery for his assistance with the photographic work; and to Dr. J. D. Cobine for assistance in procuring electrode materials.

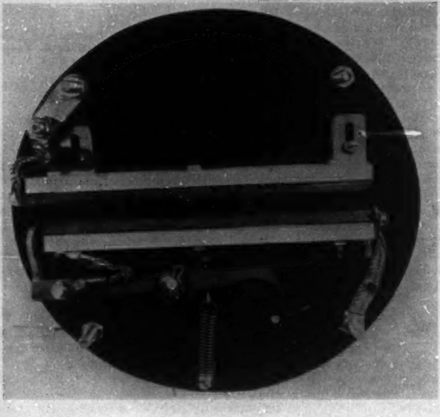


Fig. 1. Electrodes and arc-initiating mechanism

The circuit polarity was fixed so that the upper electrode was the cathode and the lower one the anode. A movable contact, connected electrically to the anode, made initial contact with the upper electrode and established the current. A cam, driven by a small synchronous motor mounted external to the solenoid, caused the movable electrode to separate from the fixed electrode to initiate the arc. The direction of the magnetic field was chosen so as to force the arc to transfer to the stationary anode and travel from left to right along the gap. The arc current was stabilized by an inductance of 40 millihenrys.

The arc was photographed periodically during transit by a still camera through the use of a simple, although effective, high-speed shutter. The shutter consisted of a revolving disk which was completely opaque except for ten transparent uniformly spaced half-degree sectors. It was made by placing a mask to duplicate the slits over a piece of photographic film and overexposing the remainder of the film. Such a shutter was very light and well balanced and posed no problems of balancing, even when turning at relatively high speeds. It was driven by an adjustable-speed d-c motor controlled by a Thymotrol drive. Fig. 2 shows a series of photographs recorded by the camera as the arc traversed the length of the electrodes. The film was subjected to a double exposure in order to include the contact assembly to orient the arc. Knowing the dimensions of the electrodes, the speed of the disk, and the spacing between slits, the distance and time between successive images could be measured and the velocity of the arc also could be calculated.

The arc voltage was recorded by a DuMont type *304A* cathode-ray oscillograph as indicated in Fig. 3. A 60-cycle 100-volt (peak-to-peak) square wave provided both a voltage and time calibration. The negative peak was adjusted so as to be coincident with the zero axis of the arc voltage. A special triggering circuit was designed and built to trigger the sweep at the instant the arc transferred from the movable to the fixed lower electrode.

Experimental Results

ARC TRAVEL

Two distinct modes of arc travel were observed. The first mode, which occurred in the lower current range, was distinguished by a cathode spot which moved continuously at a uniform velocity. The cathode spot left a continuous track, showing very little burning on the surface of the electrode. The appearance was similar to that described by previous authors for a "cold-cathode" arc.[3]

At the smallest currents, the anode spot, as well as the cathode spot, moved continuously, and the entire arc progressed along the length of the gap at a uniform velocity. Fig. 2 shows several exposures of such an arc between copper electrodes at a current of 3 amperes and flux density of 96 gauss. Because of the finite size of the slots, the shutter could not completely stop the motion so that the apparent width of the arc is somewhat greater than the true width. The arc voltage exhibited the usual random variation characteristic of the d-c arc. As might be expected, the average voltage was higher for the moving arc than for a comparable stationary arc.

When the arc current was increased to higher values, the anode spot no longer followed the uniform motion of the column, but tended to lag behind it and advance in discrete steps. Fig. 4 shows an arc between copper electrodes at a gap spacing of approximately 1/2 inch. The arc current was 71 amperes and the flux density 26 gauss. The series of photographs illustrates the manner in which the cathode and column advance under the influence of the magnetic field while the anode spot tends to lag. The formation of the new anode spots and extinction of the old ones is clearly exhibited. Apparently as the arc stretches out, the leading edge of the column is attracted toward the anode and comes close enough to the electrode to establish a new anode spot. It is not clear whether the column actually makes contact with the electrode or whether the voltage gradient simply reaches the necessary value to cause breakdown. When the new anode spot is formed, the arc tends to follow the shorter path and the previous anode spot is extinguished. During the transient period, however, multiple anode spots are often observed. The figure includes one exposure in which three such spots are in existence simultaneously.

The presence of discrete anode spots is further verified by the appearance of tracks observed on the electrode following the passage of the arc. The continuous track previously described has been replaced by a series of discrete, discontinuous spots separated by distances comparable to the spacings of the anode spots shown photographically. As might be expected from the manner in which the new anode spots are thought to be established, the separation between successive anode spots increases in general with an increase in the gap separation. The intermittent motion of the anode spot introduces a pronounced saw-tooth characteristic to the wave form of the arc voltage. Fig. 3 is a typical oscillogram, taken at a current of 12 amperes in a field strength of 29 gauss. It will be observed that the arc voltage apparently increases with time as the arc stretches out and then suddenly decreases when the formation of a new spot shortens the path.

Fig. 5 shows a family of curves of arc velocity as a function of flux density for copper electrodes with a gap spacing of 1/8 inch over a current range between 3 and 109 amperes. Owing to the fre-

Fig. 2. Photograph of a copper arc, double exposed to show electrodes; current 3 amperes, flux density 96 gauss

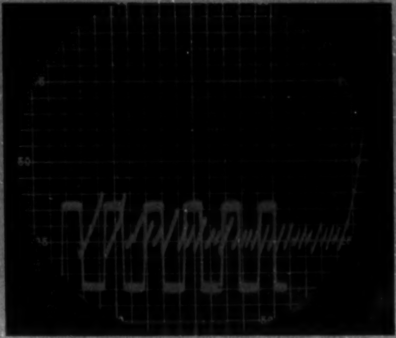


Fig. 3. Arc voltage across a copper arc; current 12 amperes, flux density 29 gauss; square-wave calibration 100 volts peak to peak, 60 cycles

quent failure of the photographs to reveal a distinct boundary to the column and to the intermittent motion of the anode spot over most of the range of the data, the velocity of the cathode spot was taken as the criterion for arc velocity. The shutter speed was so chosen that images were produced on the photograph at about 1/2-inch spacings, and the distance between spots could be measured to the nearest 1/64 inch, permitting a maximum error of measurement of the order of 3 per cent (%). The points plotted represent the average of several such velocities measured along the path of travel. In general the data follow smooth curves and are readily reproducible. The velocity along the path of travel was relatively constant for any one passage of the arc. In some measurements for copper the maximum deviation was as low as 2% of the average value while in the worst case, the maximum deviation reached 18%. With tungsten electrodes, a maximum deviation of 33% was measured.

The arc velocity was found to increase with an increase in either the arc current or flux density, and to be relatively independent of the cathode material, so long as the cathode spot advanced freely. Typical results are plotted in Figs. 5 and 6. However, when the arc current is increased beyond some critical value, dependent on both the flux density and the cathode material, the cathode spot ceases to advance with uniform motion. Rather, it tends to become fixed to one spot while the column is forced ahead by the magnetic field. When the column stretches out far enough, it may come into proximity with the negative electrode and help establish a new cathode spot. In this manner, the cathode spot tends to become oriented in discrete spots and to advance in a series of intermittent discontinuous steps. With this second mode of motion, both the cathode and the anode spots jump ahead randomly and apparently independently of each other, so that either spot may lead or lag the other, although both lag the column. The average velocity is reduced to a fraction of the velocity observed in the first mode.

The motion of the arc in the second mode of travel is illustrated by Fig. 7. The arc is between carbon electrodes at a current of 30 amperes in a magnetic field of 50 gauss. In the last four exposures of the arc, the cathode spot remained fixed in position while the anode spot continued to advance. In the first of the four exposures, the cathode spot (upper electrode) had jumped ahead of the anode spot which appeared to be in essentially the same spot as in the previous exposure. In the next exposure, the anode spot had jumped ahead of the cathode spot. In the third exposure, the anode spot had again jumped ahead while in the last exposure it had made little more progress. The failure of the cathode spot to advance further caused the arc to stretch itself out until it was extinguished. From the evidence of this illustration, it would appear that the arc column was blowing out toward the observer rather than along the electrodes so that ionization from the column could not help to establish a new cathode spot. Hence the arc was blown out before completing its travel along the electrodes.

Results for the second mode of travel are not reproducible as in the first mode, and quantitative data are limited. The arc voltage fluctuates between wide limits as both the cathode and anode spots move intermittently, and often reaches a high value when the cathode spot lags excessively. When the arc voltage reaches the open-circuit value without a new cathode spot being formed, the arc is extinguished. Sometimes the arc travels the entire length of the gap by establishing sucessive new cathode spots; other times it is extinguished in transit as demonstrated in Fig. 7. The conclusions

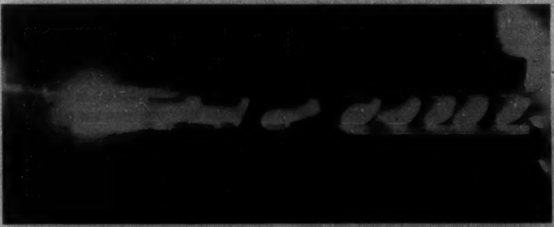


Fig. 4. Photograph of a copper arc, with top electrode cathode, bottom electrode anode; current 71 amperes, flux density 26 gauss

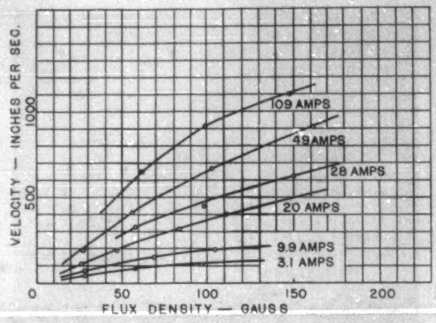


Fig. 5. Arc velocity versus flux density, copper electrodes, 1/8-inch gap

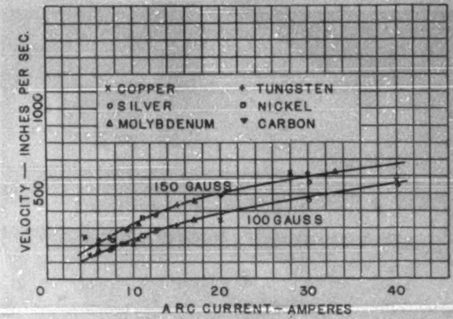


Fig. 6. Arc velocity versus current, 1/8-inch gap

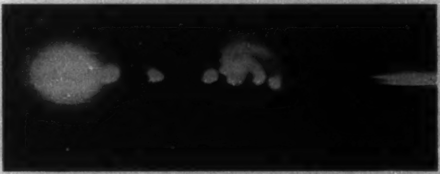


Fig. 7. Photograph of a carbon arc with top electrode cathode, bottom electrode anode, current 30 amperes, flux density 50 gauss

drawn from the photographs and oscillograms are correlated by the appearance of the cathode following arcing. The continuous tracks observed in the first mode have been replaced by discrete spots corresponding to the discontinuous cathode spots. In some cases these spots extend the length of the gap while in others they disappear part way along, verifying the incomplete passage of the arc.

The two different modes of arc travel were observed for most of the electrode materials studied. The current at which the transition was first observed was highest for copper, followed by silver. For the other materials tested, it was relatively low. The transition current increased with an increase in field strength as might be anticipated. A summary of transition currents and corresponding flux densities for a number of electrode materials has been tabulated in Table I.

Table I. Parameters for Transition in Arc Velocity

Electrode Material	Current, Amperes	Flux Density 1st Mode (High-Velocity), Gauss	 2d Mode (Low-Velocity), Gauss
Copper	109	62	25.8
	49	25.8	
Silver	55.5	100	68.8
	30	68.8	34
Tungsten	15	85	34
Molybdenum	33	155	120
Titanium	6		140
Carbon	30		155
	15	200	130
Nickel	22.5	172	172

For copper, the transition was first observed at a current of 109 amperes and a flux density of 25.8 gauss. The arc velocity was reduced to only 130 inches per second compared with about 300 inches per second predicted by extrapolating the data of Fig. 5 for an arc traveling in the first mode. A decrease in arc current to 49 amperes caused the arc to travel in the first mode again with the velocity increased to 200 inches per second. The second mode was not observed at this value of current. At 109 amperes, an increase in flux density to 62 gauss also caused the arc to revert to the first mode, increasing the arc velocity to 650 inches per second. With silver electrodes, the first mode was observed at 100 gauss for a current of 55.5 amperes, but the second mode occurred for the same current at 68.8 gauss. When the current was reduced to 30 amperes, the first mode again obtained at 68.8 gauss, but the arc reverted to the second mode again when the flux density was reduced to 34 gauss. For titanium the motion seemed to be in the second mode at currents as low as 6 amperes at a flux density as high as 140 gauss. Results for titanium were not reproducible. The second mode was observed for carbon at 30 amperes and a flux density as high as 155 gauss. When the arc current was reduced to 15 amperes, the first mode was found at a flux density of 200 gauss, but was replaced by the second mode at 130 gauss.

The transition point for nickel was found to be at a current of 22.5 amperes and a flux density of 172 gauss. Under these conditions both modes of motion were observed. When the motion was in the first mode, an arc velocity as high as 710 inches per second was measured. This speed was reduced to as low as 38 inches per second when the arc traveled in the second mode, a reduction ratio of nearly 19. Results could not be obtained with aluminum because the aluminum cathode burned easily and caused the cathode to pit, causing the column to stretch out until the arc was extinguished. This difficulty is reminiscent of experiments described by Cobine and Gallagher[4] in which the cathode spot demonstrated reluctance at crossing even a small scratch on the cathode surface.

ANODE CURRENT DENSITY

Until recently very little attention has been paid to the phenomena associated with the anode of a high-pressure high-current arc. Finkelnburg[5] has reported the high anode current density of the carbon arc and Kouwenhoven and Jones have reported the high anode current density observed in their study of an arc with a moving anode.[6] Cobine and Burger emphasized the importance of the anode region in a paper before the 1953 Gaseous Electronics Conference. A study of the tracks left by the anode of an arc moving in a magnetic field provides an excellent opportunity to estimate the current density of the anode spot. The technique employed was the same as that used by Cobine and Gallagher to measure the cathode spot current density.

A 1/2-inch gap was used during the study of the anode current density. This increased the separation between successive anode spots and helped distinguish between single and multiple spots. The tracks were observed to consist of small shiny centers surrounded by somewhat larger zones which appeared to be oxidized by the heat. The bright spots were assumed to be the anode spots. In general, they could be approximated by either circles or ellipses for the purpose of computing the area. Measurements were made for arc currents in the range between 50 and 75 amperes driven by fields of the order of 150 to 200 gauss. The anode current density was calculated as the ratio of the arc current to the area of the shiny part of the anode spot. A summary of the maximum and minimum values of the anode current density for a number of anode materials is given in Table II.

Table II. Anode Current Density

Anode Material	Anode Current Density, Amps per Sq. Cm. Maximum	Minimum
Nickel	16,000	10,000
Aluminum	36,000	17,000
Copper	44,000	26,000
Silver	44,000	28,000
Tungsten	66,000	40,000
Titanium	89,000	54,000
Molybdenum	110,000	51,000

The results show remarkable agreement with the approximation of 10,000 to 50,000 amperes per square centimeter estimated by Cobine and Burger. The higher current density measured for the highly refractory materials suggests that the anode temperature may be higher for refractory than for nonrefractory materials.

Discussion of Results

The results suggest the possibility of four types of motion for the d-c arc in a magnetic field:

1. Continuous motion of both the cathode and anode spots.
2. Continuous motion of the cathode spot and intermittent motion of the anode spot.
3. Continuous motion of the anode spot and intermittent motion of the cathode spot.
4. Intermittent or discontinuous motion of both the cathode and anode spots.

Of the four possibilities, the third was never experienced. In every case observed the motion of the anode spot became discontinuous at a lower current than the cathode spot. This is reasonable because the higher temperature normally associated with the anode would be expected to stabilize the position of the anode spot at a lower current than the cathode spot.

Continuous motion of both the anode and cathode spots was limited to low currents and high field strengths. The intermittent motion of the anode spot was observed at currents of only a few amperes and was characteristic of most of the arcs studied. No discontinuity was discovered in the magnitude of the arc velocity when the anode spot ceased to move continuously. This justifies the grouping of all arcs with a continuously moving cathode in the same mode of motion. The decreasing slopes of the curves for arc velocity as the current or flux density is increased, as illustrated by Figs. 5 and 6, suggest however that the discontinuous motion of the anode spot introduces a retarding force on the arc. The values of arc velocity for copper, silver, molybdenum, tungsten, nickel, and carbon were all found to follow the same curves, as demonstrated by Fig. 6, suggesting the conclusion that the arc velocity may be independent of the electrode material so long as the arc follows the first mode of motion. Stolt's conclusion that the arc velocity was independent of the cathode material suggests that his results were all obtained for arcs traveling in the first mode of motion. Quite likely, if he had increased the current with carbon electrodes, he would have experienced the effect of the second mode of motion on the arc velocity.

The discontinuous motion of the anode spot shows striking similarity to the motion of the anode spot of an arc between a fixed cathode and a moving anode, as observed by Kouwenhoven and Jones.[6] In their work, performed in the absence of a magnetic field, the anode spot tended to become fixed to the workpiece (anode) and was carried along, stretching out the column in much the same way as observed in this study, until a new spot was established on the workpiece opposite the cathode. In the present study, the anode spot again tended to hang onto the same spot on the anode surface while the magnetic field forced the cathode spot and column ahead. The general appearance of the column and the mechanism of establishing new anode spots appear to be similar in the two cases.

High peaks were left by the anode spots, similar to those described by Kouwenhoven and Jones. The peaks were highest with copper and relatively less pronounced for refractory materials. At low currents and high magnetic fields, the anode sometimes exhibited a continuous track or a track broken at a few spots. The formation of the peak is believed to be independent of the motion of the anode since a similar peak has been observed on the silver anode of an a-c contactor following arc extinction without the aid of a blowout coil. Alternating current was used and the arc current was interrupted at the normal current zero. The contact area was small and no evidence of motion of the anode spot was observed.

The discovery of the second mode of motion for the cathode is of basic significance to both the fundamental scientist and the design engineer. To the one, it adds to the knowledge of the fundamental behavior of the electric arc; to the other, it places a limitation on the amount of arc current that can be driven by a magnetic field. The transition in effect places a ceiling on the curent that can be driven at high speed in a reproducible manner. The ceiling can in general be kept at a maximum by the choice of a nonrefractory cathode and raised by an increase in the field strength. The transition current at which the cathode spot ceases to move continuously appears to be determined by the cathode material only. Measurements made with a copper or silver cathode and a refractory anode produced arc velocities in the first mode of motion for arc currents well above the transition values for cathodes of the same material as the anode. In this way, anode current densities (Table II) were measured with the arc traveling in the first mode for all the anode materials at currents far above the transition values that would have been encountered had both electrodes been of the same composition.

A new cathode spot is much more difficult to establish than a new anode spot so that reproducible results could not be obtained for the second mode of motion. With a clean electrode, the cathode spot tended to stick even at a low current and high field and irregular results were experienced for the first few passages of the arc with all electrode materials. Even for copper and silver cathodes, the cathode and anode spot both tended to stick during the first few operations and then the cathode spot began to move with discontinuous motion. Sometimes the cathode spot would move part way along the gap and then become extinguished. After several operations in this mode, the first mode would appear and would be the normal mode. When the cathode was highly oxidized at the start, the motion would initially be in the first mode. The need for an oxidized cathode has been recognized many times as essential to the establishment or maintenance of a cathode spot.

In most of the tests, the magnetic field was excited before the arc was initiated. When the arc was established first, the cathode tended to establish a hot spot before the arc started to move and the cathode would stick as if it were in the second mode. If the arc current and flux density were in the range that would normally permit motion in the first mode, however, the arc would start to move in the first mode as soon as a new cathode spot was formed. Even when the magnetic field was excited first, considerable delay occurred between the initiation of the arc and the start of the motion. This time, referred to as "sticking time" in contactor applications, was often several times greater than that which was required for the complete passage of the arc along the gap.

References

1. Rotation of the Electric Arc, **H. Stolt.** *Annalen der Physik*, Leipzig, Germany, vol. 7, no. 1, Apr. 1924, pp. 80–104.

2. The Retrograde Motion of the Arc Cathode Spot, **C. J. Gallagher.** *Journal of Applied Physics*, American Institute of Physics, New York, N. Y., vol. 21, no. 8, Aug. 1950, pp. 768–71.

3. Theory of the Deion Circuit-Breaker, **J. Slepian** *AIEE Transactions*, vol. 28, 1929, pp. 523–27.

4. Current Density of the Arc Cathode Spot, **J D. Cobine, C. J. Gallagher.** *Physical Review*, vol. 74, no. 10, Nov. 1948, pp. 1524–30.

5. The High Current Carbon Arc, **W. Finkelnburg.** *Fiat Final Report No. 1052*, Office of Military Government for Germany, U. S. Department of Commerce, Washington, D. C., 1946.

6. Characteristics of Arcs Between Moving Electrodes, **W. B. Kouwenhoven, T. B. Jones.** *Electrical Engineering*, vol. 68, Oct. 1949, p. 834.

◆

Discussion

C. G. Smith (Raytheon Manufacturing Company, Newton, Mass.): The magnetic field transverse to the arc current forces the air to one side in a blast that can be observed by suitably arranged vanes. The blast is in the expected direction and is probably the main factor causing the arc motion. However, there is another factor at work tending to move the arc in the opposite direction. This is revealed if one minimizes the air blast.

The electromagnetic force at given field strength is proportioned to the arc length times the arc current. The viscous drag on the air is inversely proportioned to arc length or spacing between terminals. Closer spacing of the terminals reduces the impelling force and increases the retarding force, thus minimizing the air blast. Some years ago I observed a copper arc between terminals about 0.030 inch apart at a variable pressure in the region of 10 centimeters of Hg. At the low pressures and with close spacing, the arc motion was retrograde (opposite to the expected direction). The vanes for detecting the blast of air always moved in the proper or expected direction. The vanes therefore moved contrary to the observed arc motion when the arc was proceeding in the retrograde direction.

It would be of interest if the observations could be repeated or extended by employing terminals whose spacing could be made quite small, the gas pressure being under control. The phenomena are conceivably of practical interest in deion circuit breakers or the like employed at high elevations with attendant reduced pressure. Furthermore the basic problem of arc motion in a transverse field is of scientific interest because we seem not to understand the retrograde motion; and we probably do not understand the sticking to a localized anode or cathode region as observed.

L. P. Winsor: The authors appreciate learning of Dr. Smith's observations concerning the retrograde motion of the arc in a magnetic field at low pressures. Their own experiments were conducted at atmospheric pressure and with a somewhat longer gap, so that the retrograde motion was not in evidence. However, there is no doubt that an extension of the work to lower pressures might lead to useful results both from a theoretical and from a practical viewpoint.

Vol.
Pt. 1
1956

Magnetic Effects of Compressional Stress at Low Field Intensities

R. E. FISCHELL
ASSOCIATE MEMBER AIEE

Synopsis: Toroidal, laminated cores of several important magnetic materials were tested for the low-field-intensity magnetic properties at compressional stresses ranging from 0 to 500 pounds per square inch (psi). The compressional stress was applied in a direction perpendicular to the plane of the laminations and perpendicular to the direction of the magnetic field, thereby simulating the binding of magnetic laminations in transformers, electric motors, servo components, etc. The materials tested were hydrogen-annealed specimens of 3.5 per cent (%) and 6.4% silicon-iron, 81% and 90% nickel-iron, *4-79* and *5-79* molypermalloy, and hydrogen- and air-annealed specimens of *AEM-4750*, *12* Alfenol, and *16* Alfenol.

Two types of tests were made. The first test determined the effect of compressional stress on the d-c permeability for a range of magnetic field intensities from 0.1 to 100 milliorsteds. The second test determined the effect of compressional stress on the 60-cycle per second (cps) a-c magnetization curve for a range of peak magnetizing fields from 0.01 to 100 milliorsteds. Under the action of an increasing compressional stress, each of the materials tested exhibited a decrease in low field permeability and an a-c magnetization curve that was shifted downward. The degree to which the permeability and a-c flux density changed with increasing compressional stress varied widely from one material to another, but for each material the d-c permeability and a-c flux density decreased by approximately the same percentage of the zero pressure measurement of these quantities.

MAGNETIC materials are very frequently used in the form of a clamped stack of laminations. For this configuration the direction of the clamping stress is perpendicular to the plane of the laminations and perpendicular to the direction of the applied magnetic field. It is therefore of interest to observe the

---

Paper 56-133, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted July 8, 1955; made available for printing December 30, 1955.

R. E. Fischell is with the Naval Ordnance Laboratory, Silver Spring, Md.

The author gratefully acknowledges the assistance of R. H. Lundsten for his aid in the performance of the experiments and of D. I. Gordon for his helpful suggestions and criticisms in the preparation of the manuscript.

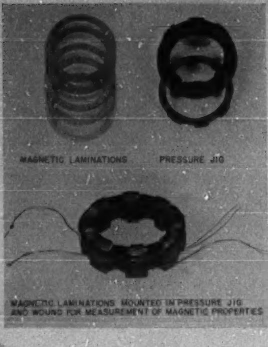

MAGNETIC LAMINATIONS
PRESSURE JIG
MAGNETIC LAMINATIONS MOUNTED IN PRESSURE JIG AND WOUND FOR MEASUREMENT OF MAGNETIC PROPERTIES

Fig. 1. Special jig for applying a transverse compressional stress to magnetic laminations

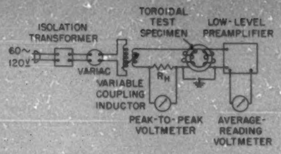

ISOLATION TRANSFORMER
60~ 120V
VARIAC
TOROIDAL TEST SPECIMEN
LOW-LEVEL PREAMPLIFIER
VARIABLE COUPLING INDUCTOR
R_H
PEAK-TO-PEAK VOLTMETER
AVERAGE-READING VOLTMETER

Fig. 3. Circuit for measuring a-c flux densities at low magnetizing fields

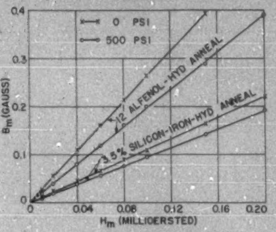


Fig. 4. A-c magnetization curves at 60 cps for 12 Alfenol and 3.5% silicon-iron

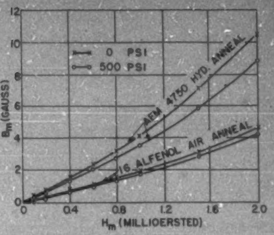


Fig. 5. A-c magnetization curves at 60 cps for AEM-4750 and 16 Alfenol

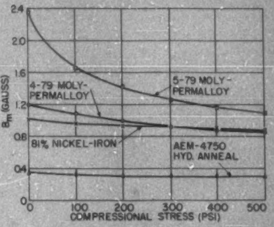


Fig. 6. A-c flux density as a function of compressional stress for AEM-4750, 81% nickel-iron, 4-79 and 5-79 moly-permalloy

effect of this transverse compressional stress on the magnetic properties of laminations.

A previous paper[1] discusses the effect of transverse compressional stress on the a-c hysteresis loop and magnetization curve, and on the maximum permeability, coercive force, remanence, and saturation flux density of several important magnetics materials. The purpose of this paper is to describe the effect of transverse compressional stress on the a-c and d-c magnetic properties at very low magnetic field intensities.

The materials tested were hydrogen-annealed specimens of 3.5% and 6.4% silicon-iron, 81% and 90% nickel-iron, *4-79* and *5-79* moly-permalloy, and hydrogen- and air-annealed specimens of *AEM-4750*, *12* Alfenol, and *16* Alfenol. The specimens tested were laminated toroids with thin-paper insulation between the laminations. The toroids measured 1⁵/₈-inch inside diameter and 2.0-inch outside diameter; the lamination and insulating paper thicknesses measured 0.015 and 0.001 inch respectively.

Experimental Procedure

The compressional stress was applied by placing the laminations in a special fixture, shown in Fig. 1. With the use of this device, a compressional stress of varying magnitude was exerted on the sample without interfering with the electrical windings which are necessary for the determination of magnetic properties.

The device for applying the compressional force was a lever and scale system, shown in Fig. 2. With this system it was possible to maintain constant pressure on the sample while it was being tested, and it was also possible to increase the compressional stress in increments of 100 psi without relieving the stress while changing from one pressure to another.

Two types of tests were made, and for each test the magnetic properties were measured at compressional stresses of 0, 100, 200, 300, 400, and 500 psi, and the low field properties were also measured after the stress had been removed. The first of these tests was the measurement of maximum a-c flux density at a frequency of 60 cps with peak magnetizing fields ranging from 0.01 to 100 millioersteds. The second test was the measurement of d-c permeability at magnetic fields ranging from 0.1 to 100 millioersteds.

The circuit for the a-c measurements is shown in Fig. 3. The magnetizing current was controlled by means of a Variac and a variable-coupling air-core mutual inductor. With this arrange-

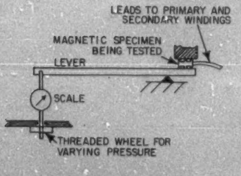

LEADS TO PRIMARY AND SECONDARY WINDINGS
MAGNETIC SPECIMEN BEING TESTED
LEVER
SCALE
THREADED WHEEL FOR VARYING PRESSURE

Fig. 2. Lever and scale system for applying a transverse compressional stress to magnetic laminations

Table I. Percentage Decrease in A-C Flux Density at H_m=0.10 Millioersted

Material	Anneal	Compressional Stress, Psi 100	200	300	400	500	O_R
12 Alfenol	hydrogen	8.33	14.29	18.45	22.02	24.40	8.93
12 Alfenol	air	1.15	1.92	2.31	2.88	3.65	0
16 Alfenol	hydrogen	0.57	4.66	5.01	6.94	11.83	
16 Alfenol	air	1.15	2.43	3.58	5.24	8.17	0.13
AEM-4750	hydrogen	5.79	7.72	9.27	9.65	11.20	1.35*
AEM-4750	air	2.70	0.90	5.41	5.41	5.86	
3.5% silicon-iron	hydrogen	3.42	6.62	8.68	10.73	13.01	0.46
6.4% silicon-iron	hydrogen	1.77	3.54	4.42	5.31	5.31	4.42*
81% nickel-iron	hydrogen	10.67	13.11	14.67	16.67	18.00	2.67
90% nickel-iron	hydrogen	17.47	25.70	32.73	36.95	43.17	11.85
4-79 moly-permalloy	hydrogen	8.77	17.54	23.39	26.32	30.41	0.58
5-79 moly-permalloy	hydrogen	29.41	39.93	46.41	50.39	53.46	8.50

* Percentage increase.

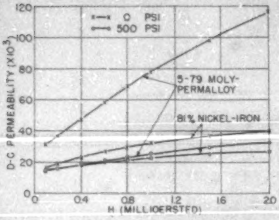


Fig. 7. D-c permeability at low field intensities for 81% nickel-iron and 5-79 moly-permalloy

Table II. Percentage Decrease in D-C Permeability at H=1.0 Milloersted

Material	Anneal	Compressional Stress, Psi					
		100	200	300	400	500	0_R
12 Alfenol	hydrogen	3.45*	5.17	10.34	15.52	22.41	3.45*
12 Alfenol	air	0	4.55	9.10	13.64	27.27	
16 Alfenol	hydrogen	2.22	6.67	8.89	11.11	15.56	11.11
16 Alfenol	air	2.14†	3.21†	4.70†	5.34†	5.98†	1.07†
AEM-4750	hydrogen	1.15	4.21	11.88	14.18	16.09	0.38
AEM-4750	air	11.00*	11.00	11.00	15.00	11.00	
3.5% silicon-iron	hydrogen	0	0	6.25	6.25	6.25	6.25*
6.4% silicon-iron	hydrogen	5.41	5.41	5.41	5.41	5.41	5.41
81% nickel-iron	hydrogen	12.87	19.55	23.27	26.49	28.71	8.66
90% nickel-iron	hydrogen	10.42	20.83	31.25	35.42	39.58	10.42
4-79 moly-permalloy	hydrogen	6.74	12.95	21.76	24.87	32.64	1.04
5-79 moly-permalloy	hydrogen	40.66	53.85	57.14	60.44	64.84	10.99

* Percentage increase.
† Measurement made at 6.0 millioersted.

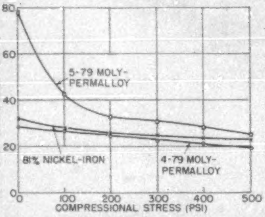


Fig. 8. D-c permeability as a function of compressional stress for 81% nickel-iron, 4-79 and 5-79 moly-permalloy

ment it was possible to obtain and control the very low voltages which were necessary to produce the extremely small magnetizing currents required for the a-c measurements. In terms of the peak-to-peak voltage measured across the primary current metering resistance R_H, the peak magnetizing field is given by the relation

$$H_m=\frac{0.2\pi N_1}{L_m R_H}V_{p-p} \text{ oersteds} \tag{1}$$

where

N_1 = number of primary turns
R_H = metering resistance in ohms
L_m = mean length of the toroid in centimeters
V_{p-p} = peak-to-peak voltage in volts

In the a-c experiments, 32 evenly spaced turns were used for the primary winding, and with the use of a metering resistance of 139.0 ohms the relation between the peak-to-peak voltage and the magnetizing field takes the simple form

$$H_m=10V_{p-p} \text{ millioersteds} \tag{2}$$

To measure the peak a-c flux density produced at a very small value of H_m it was necessary to use a low-level preamplifier in conjunction with a sensitive electronic vacuum tube voltmeter of the rectified average-reading type. With the circuit shown in Fig. 3, the combined noise level of the preamplifier and voltmeter was equivalent to an input signal of less than 2.5 microvolts. This noise level was considerably below the lowest signal measured in the a-c tests.

In terms of the readings of a rectified average-reading voltmeter the maximum a-c flux density is given by

$$B_m=\frac{V_{2_{\text{rms}}}\times 10^8}{\pi\sqrt{2}mN_2Af} \text{ gauss} \tag{3}$$

where

$V_{2_{\text{rms}}}$ = voltage as measured by a rectified average-reading voltmeter calibrated in rms voltage of a sine wave
m = gain of the preamplifier
N_2 = number of secondary turns
A = cross-sectional area of the toroid in square centimeters
f = frequency in cps

The initial permeability can be determined from a-c measurements with the use of the relation

$$\mu_0=\lim_{H_m\to 0}\frac{B_m}{H_m} \tag{4}$$

$$\mu_0=\lim_{H_m\to 0}\frac{L_m R_H V_{2_{\text{rms}}}\times 10^8}{0.2\sqrt{2}\pi^2 mN_1N_2AfV_{p-p}} \tag{5}$$

The second test used the standard d-c fluxmeter circuit to determine the permeability at low values of magnetic field. For the d-c case, the magnetizing field is given by the familiar toroid equation

$$H=\frac{0.4\pi N_1 I_1}{L_m} \text{ oersted} \tag{6}$$

where I_1 = primary current in amperes.

The flux density can be found from the relation

$$B=\frac{K}{2N_2A}d \text{ gauss} \tag{7}$$

where

K = fluxmeter constant in maxwell-turns per centimeter
N_2 = number of secondary turns
d = fluxmeter deflection in centimeters

The permeability will then be given by

$$\mu=\frac{B}{H}=\frac{L_mKd}{0.8\pi N_1N_2AI_1} \tag{8}$$

and the initial permeability is given by

$$\mu_0=\lim_{H\to 0}\frac{L_mKd}{0.8\pi N_1N_2AI_1} \tag{9}$$

The value of the initial permeability obtained from equation 5 with a-c measurements should be equivalent to the value obtained in equation 9 with d-c data.

Magnetic measurements at extremely low field levels require precautions which can usually be neglected when most magnetic properties are measured. Three important factors which had to be considered in the performance of these low-field experiments will be discussed briefly.

1. The effect of an external field large compared to the impressed magnetizing field.
2. Demagnetization of the specimens.
3. Means of eliminating or minimizing electromagnetic pickup.

Since measurements were made at fields of less than 1/10,000 that of the earth's field, it was necessary to determine if a comparatively large external field can cause errors in the measurement of flux density. To determine the effect of a comparatively large external field, the low-level flux density was measured at three different magnitudes of external field: 1. less than 10^{-5} oersteds, 2. the earth's field, and 3. double the earth's field. The results of these experiments were negative, i.e., the apparent flux density of the toroid was unaffected by external fields of these magnitudes. It was therefore decided that the experiments could be accurately performed without shielding the toroids from the earth's magnetic field.

When demagnetizing to very low flux levels it is important to use a smoothly decreasing magnetic current such as that produced by a variable-coupling mutual inductor. The small but discrete voltage steps produced by turning down a device such as a Variac can sometimes cause the material to be left in an undemagnetized condition.

For the a-c measurements, stray electromagnetic pickup must be minimized by grounding the system to a quiet ground, as shown in Fig. 3, and by minimizing the resistance of the secondary winding.

Experimental Results

Some typical results of the effect of compressional stress on low-field a-c magnetic properties are shown in Figs. 4, 5, and 6. Fig. 4 shows the effect of exerting a compressional stress of 500 psi on the a-c flux density of hydrogen-annealed specimens of *12* Alfenol and 3.5% silicon-iron. It should be noted that these curves are straight lines for these extremely low field intensities, and it can therefore be assumed that the region of initial permeability is reached, as defined in equation 4. The compressional stress decreased the value of the initial permeability of all the materials that were tested. This decrease in initial permeability at higher values of stress is seen in Fig. 4 as a decrease in the slope of the straight-line magnetization curve.

Fig. 5 shows the effect of a 500-psi stress on two typical materials, air-annealed *16* Alfenol and hydrogen-annealed *AEM-4750*. The field intensities used for Fig. 5 are one order of magnitude greater than those shown in Fig. 4. From the nonlinear behavior at the ends of the magnetization curves shown in Fig. 5, we see that these materials are leaving their region of initial permeability. At these higher fields, the effect of the compressional stress was to shift the magnetization downward without an appreciable change in its shape.

Fig. 6 shows the effect of increasing the compressional stress while keeping the value of the peak magnetizing field constant at $H_m = 0.10$ millioersteds. The materials shown here are *4-79* and *5-79* moly-permalloy, 81% nickel-iron, and *AEM-4750*. These curves show that the *5-79* moly-permalloy had the highest value of initial permeability and also the greatest sensitivity to compressional stress.

The results of the a-c measurements on all the materials tested are shown in Table I. The numbers in the column headed O_R are the percentage change in B_m after the stress was removed. Most of the materials returned within a few per cent to their prestressed value of B_m.

Some typical results of the d-c measurements of low field permeability as a function of compressional stress are shown in Figs. 7 and 8. The d-c permeability as a function of field strength is plotted in Fig. 7 for the materials *5-79* moly-permalloy and 81% nickel-iron at compressional stresses of 0 and 500 psi. The higher stress decreased the permeability at all values of the magnetic field. This was true for all the materials tested. Fig. 8 shows the result of increasing the compressional stress on specimens of *4-79* and *5-79* moly-permalloy and 81% nickel-iron while holding the field intensity constant at $H = 1.0$ millioersted. These curves are strikingly similar in shape to the a-c flux density curves as a function of stress, as shown in Fig. 6.

The results of the d-c permeability measurements on all the materials tested are shown in Table II.

Conclusions

From the experimental results, the following conclusions regarding the effect of transverse compressional stress at low field intensities can be drawn:

1. Increasing the compressional stress causes the a-c magnetization curve to be shifted downward. At very low fields this is equivalent to a decrease in the initial permeability.

2. Increasing the compressional stress decreases the low field intensity d-c measurement of permeability.
3. The percentage change in magnetic properties varies widely from one material to another but, in general, materials with the highest initial permeability showed the greatest stress sensitivity.
4. At magnetic fields over one oersted the compressional stress sensitivity was found to be dependent on the magnitude of the magnetostrictive constant of the material.[1] At low fields, however, the variation of magnetic properties with compressional stress does not appear to have any correlation with magnetostriction.

Reference

1. EFFECTS OF TRANSVERSE COMPRESSIONAL STRESS ON MAGNETIC LAMINATIONS, R. E. Fischell. *AIEE Transactions*, vol. 74, May 1955, pp. 190–93.

Discussion

John Cybulski (Naval Research Laboratory, Washington, D. C.): Would the author comment on whether the effects of this stress may be explained qualitatively by considering that a normal compressive stress on an oblique plane can be resolved into a compressive and tangential (shearing) component? The maximum value of this shearing stress occurs on a plane making a 45-degree angle with the plane on which the normal stress occurs and is one-half that value. In general, the shearing strength is much less than the compressive or tensile strength so that it may be the more significant stress. This stress may then be along one of the various preferred axes of orientation of the ferromagnetic crystalline structure, e.g., the 111 direction where 001 is taken to be in the direction of rolling and the 110 direction is perpendicular to it and in the plane of the sheet.

The effects of simple tensile and compressive stresses in the direction of the field on the magnetic characteristics has been covered previously. However, Poisson's ratio relating the lateral dimensional change to the longitudinal change (about one-fourth for steel) may be also significant under transverse stresses.

It would be interesting to determine the effects of hydrostatic pressure on these materials to see if there would be a reinforcement of these shearing stresses.

The possible lack of correlation between the magnetostrictive properties and the effects of transverse compressional stress at low field intensities may be due only to the fact that at present there are insufficient data available on the magnetostriction constant at the low intensities studied. In the author's first paper (see reference 1 of the paper) there was some correlation which might be related to the shearing stresses discussed. Many materials, notably, the iron-silicon alloys, exhibit negative and positive magnetostrictive properties, depending on the field intensities supported by the core material.

R. E. Fischell: No material used for these experiments on transverse compressional stress was grain-oriented. Since these are polycrystalline materials having no preferred axis of orientation, there is no reason to believe that the observed decrease in magnetization is due merely to the presence of a shearing stress along an oblique plane. Although the lateral dimensions are changed according to Poisson's ratio, no qualitative explanation can be found here because this same strain was present for materials having both positive and negative magnetostriction, and yet each showed a decrease in magnetization with increasing stress.

Experiments now being performed at the Naval Ordnance Laboratory to determine the effect of hydrostatic pressure on magnetic materials may lead to a better understanding of the effect of transverse stress on magnetization. I agree that the phenomenon at low field intensities may have some correlation with magnetostriction, but at present there are no data known to the author on magnetostriction at magnetic field intensities below 1.0 millioersted.

A Digital Computer Program for Determining Atomic-Wave Functions

W. W. PIPER
NONMEMBER AIEE

SOLUTION of a set of second-order differential equations for computing Hartree-Fock self-consistent field wave functions has been successfully programmed for use on an International Business Machines Corporation (IBM) type *650* magnetic drum digital computer. These equations are nonhomogeneous and have coefficients which are functions of the solution. Thus, the equations must be integrated a sufficient number of times to reach a state of self-consistency. For a particular atomic configuration this may be accomplished in 5 to 8 hours of machine time with the present program. This represents a tremendous saving over hand methods previously used.

Accurate atomic-wave functions are of considerable use and importance in certain branches of solid-state physics. The exact wave functions are solutions of the Schrödinger equation, a differential equation which is basic to quantum mechanics. For 1-electron atomic systems, such as H (hydrogen), He^+ (helium), and Li^{+2} (lithium), the Schrödinger equation can be solved exactly and the resulting wave function can be written down in analytical form. If a system consists of two or more electrons, there no longer exists an analytic solution and an approximation must be sought. The Hartree-Fock equations yield wave functions which represent most accurately many of the properties of an atomic configuration. They were first conceived over 20 years ago by Hartree[1] and Fock.[2] These equations are a set of interconnected second-order differential equations which are restricted by certain boundary conditions as well as conditions of orthonormality. The excellence of these equations is handicapped by the lengthy and laborious numerical work that is required to arrive at a solution for a particular configuration. For this reason, only 24 or so computations have been carried through, by use of desk calculating machines in the past 20 years. Depending on the experience of the person carrying out the calculations, a few months to a year is required to determine a solution. With the advent of high-speed digital computers, these solutions can be determined in a few hours.

The fundamental investigation of luminescence processes serves as a good example of the type of work for which these atomic-wave functions are useful. Luminescent materials are commonly known as phosphors. Well known examples of phosphors are the luminescent components of cathode-ray tubes and fluorescent lamps. The spectral characteristics of emission and absorption of light by a phosphor are determined by the activator impurity incorporated in the phosphor. The activator material is atomically dispersed throughout the crystalline host lattice with a specific host lattice environment. An example is potassium chloride activated with thallium. A model of this system is shown in Fig. 1. Each thallium activator center is located at a potassium ion site and may be treated as surrounded by an infinite lattice of potassium chloride. This system has been treated theoretically from fundamental principles by Williams,[3] and is the first phosphor system to be so calculated. The calculated emission and absorption bands are in excellent agreement with experimental data. In order to calculate these spectra, the total energy of the lattice must be calculated with the thallium ion in the ground state and in the lowest excited state. In order to calculate some of the terms contributing to the energy of the lattice, it is necessary to know the radial charge distribution of the electron cloud surrounding the ion, that is, the charge density at a given radius from the nucleus. It is this information that can be best approximated by a Hartree-Fock calculation.

In the past, specific phosphor systems have been developed in an empirical manner. However, due to the pioneering work of Williams, and other recent advances in the understanding of the underlying physics of impurity systems in crystalline solids, the interpretation of phosphor phenomena is rapidly becoming intelligible from fundamental considerations. As the solid-state physics of a variety of phosphors becomes evident in increasing detail, the task of optimizing these and new phosphors for specific jobs will certainly become correspondingly simplified. An important link in such a program would be the ready availability of accurate atomic-wave functions such as the Hartree-Fock functions. The use of a digital computer such as a IBM *650* has now made it possible to compute several such wave functions in a reasonable length of time.

Numerical Procedure

The Hartree-Fock scheme consists mathematically of a finite set of second-order nonhomogeneous linear differential equations with coefficients which are integral functions of the solutions of all the equations. The solutions can be approximated to the accuracy of the difference equations by an iterative technique.

There is a wave function $P_\alpha(r)$ associated with each shell of electrons (electrons with the same principle and orbital angular momentum quantum number). For example, the ground state of Cu^+ (copper) has six shells so that there are six functions $P_\alpha(r)$ and also six interconnected differential equations. For an ion with n shells,

$$
\begin{aligned}
&\frac{d^2}{dr^2}P_1(r)+g_1(r)P_1(r)+f_1(r)=0\\
&\quad\cdots\cdots\\
&\frac{d^2}{dr^2}P_\alpha(r)+g_\alpha(r)P_\alpha(r)+f_\alpha(r)=0 \qquad (1)\\
&\quad\cdots\cdots\\
&\frac{d^2}{dr^2}P_n(r)+g_n(r)P_n(r)+f_n(r)=0
\end{aligned}
$$

$P_\alpha^2(r)dr$ represents the amount of charge contained in a volume between spheres of radius r and $(r+dr)$ due to one electron in the αth shell. Since $P_\alpha^2(r)$ describes the radial distribution of one electron in all space, this function must be normalized.

$$\int_0^\infty P_\alpha^2(r)dr=1 \qquad (2)$$

In addition, this function must be orthogonal to all the other wave functions P_α

Paper 56-155, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing December 22, 1955.

W. W. PIPER is with the General Electric Company, Schenectady, N. Y.

The author expresses appreciation to Prof. D. R. Hartree for several invaluable discussions during the early stages of this work, and to Miss Mary Tressel and Miss Elise Krieger for their help in coding and correcting the program and for running the problem on the IBM *650*.

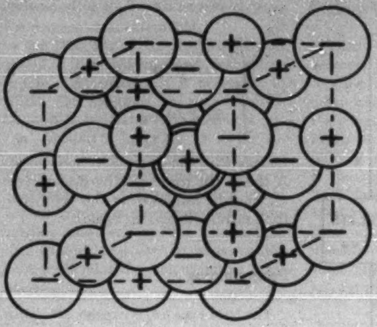


Fig. 1. KCl–Tl which is a representative example of an inorganic phosphor system. The ion at the center of the cube is a thallium ion substituted at a potassium ion site. The other positive spheres represent potassium ions and the negative ones chlorine

with the same orbital angular momentum, l,

$$\int_0^\infty P_\alpha P_\beta dr = \delta_{\alpha\beta},\ (l_\alpha = l_\beta) \tag{3}$$

and it must disappear at the boundaries $r = 0$ and $r = \infty$,

$$P_\alpha(0) = P_\alpha(\infty) = 0 \tag{4}$$

The functions g_α and f_α of equation 1 which are associated with the equation for a particular shell, depend on the wave function P_α and also on all the other wave functions. For this reason the solution P_α of each of the shells must in turn be repeated with better estimates of g_α and f_α until a state of self-consistency has been reached. That is, until a set of solutions of equation 1 agree with the functions that were used to compute the g_α's and f_α's. These coefficients are of the form

$$g_\alpha = \frac{\gamma_\alpha}{r} - \frac{L_\alpha}{r^2} - \epsilon_{\alpha\alpha} \tag{5}$$

$$f_\alpha = \frac{\phi_\alpha}{r} - \sum_\beta^{r}{}' \epsilon_{\alpha\beta} P_\beta \tag{6}$$

where

$$\gamma_\alpha = N - \sum_\nu b_\nu Y_{0\nu\nu} + \sum_i c_i Y_{n_i \alpha\alpha} \tag{7}$$

$$\phi_\alpha = \sum_k d_k Y_{n_k \alpha\beta_k} P_{\beta_k} \tag{8}$$

and

$$Y_{n\alpha\beta}(\rho) = \int_0^\rho P_\alpha(r) P_\beta(r) \left(\frac{r}{\rho}\right)^n dr + \int_0^\infty P_\alpha(r) P_\beta(r) \left(\frac{\rho}{r}\right)^{n+1} dr \tag{9}$$

Most of (γ_α/r) is the coulomb potential experienced by an electron in the αth shell, (L_α/r^2) is the effective potential due to the angular momentum of the αth shell, and $\epsilon_{\alpha\alpha}$ is related to the total energy of the αth shell. ϕ_α/r is an exchange potential which represents noncoulombic interactions between the shells. $\epsilon_{\alpha\beta}$ is adjusted so that the condition expressed in equation 3 may be satisfied. $\epsilon_{\alpha\alpha}$ is varied to satisfy conditions in equations 2 and 4. The other variable at one's disposal to help satisfy these two conditions is the slope of the function at the origin. There exist many solutions for P_α in equation 1 which satisfies the conditions of equations 2, 3, and 4. However, each solution has a different number of nodes, and the solution with the proper number of nodes may be chosen from quantum mechanical considerations.

Integration of the differential equation 1 and execution of the quadratures in equation 9 cannot be carried out analytically so that approximate solutions must be computed by numerical techniques. The most convenient schedule for many numerical procedures is to adopt a set of equally spaced points along the axis of the independent variable. For the problem discussed here, this arrangement is not suitable. Instead, the independent variable, r, is divided into groups of an even number of equally spaced intervals. The interval length changes by a factor of 2 between adjacent groups. This arrangement produces a minimum of complications that must be handled at a change of interval size. Fig. 2 illustrates the subscript convention used in the rest of this section. The value of r at the end of an interval is denoted by a subscript and the value of a function at that point by the same subscript. The subscripts increase by one with increasing values of r. The interval length is denoted by h. Thus,

$$r_{i+1} = r_i + h$$

$$f_\alpha(r_i) = f_i \tag{10}$$

Numerical quadratures are approximated by fitting a parabola through the three points associated with a pair of intervals and calculating the area under this parabola. This area may be expressed analytically and will be reduced to

$$\int_{r_i}^{r_{i+1}} f(r)dr = \frac{h}{12}(5f_i + 8f_{i+1} - f_{i+2}) \tag{11A}$$

$$\int_{r_{i+1}}^{r_{i+2}} f(r)dr = \frac{h}{12}(-f_i + 8f_{i+1} + 5f_{i+2}) \tag{11B}$$

These expressions have the advantage of not requiring any interpolation of values near the point where a change of interval size occurs. For an even number of intervals, equation 11 reduces to Simpson's rule. The error which is introduced by this approximation is approximately $(\delta^4 P/10)$.

The integration of any one of the equations in equation 1 subject to the boundary condition equations 2, 3, and 4, has been divided into three differential equations for machine computations according to a scheme suggested by Hartree in a private communication. Each equation is integrated by a difference equation technique which is discussed by Milne.[4] This difference equation can be used to obtain an approximate solution of a linear second-order differential equation at a set of equally spaced points. For an equation of the form of equation 1, the solution at r_{i+1} is given by

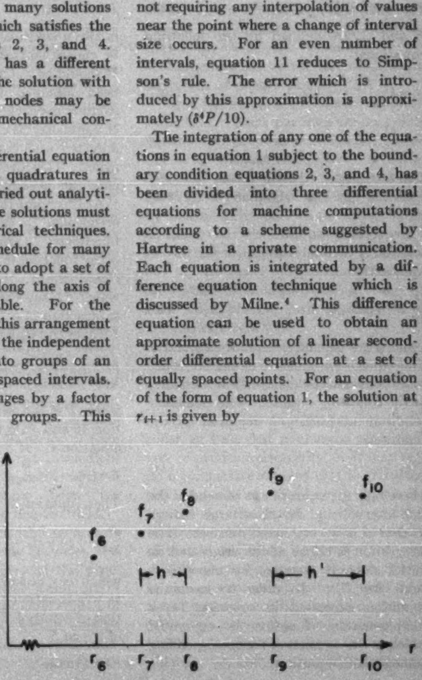


Fig. 2. Illustration of the subscript notation for a function, f(r), represented at a finite number of equi-spaced values of r. A change of interval size is also shown

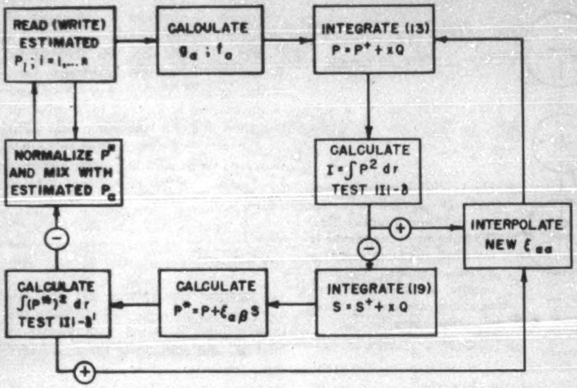

READ (WRITE) ESTIMATED P_i; $i = 1, \dots n$
CALOULATE g_a ; f_a
INTEGRATE (13) $P = P^+ + xQ$
NORMALIZE P^* AND MIX WITH ESTIMATED P_a
CALCULATE $I = \int P^2\,dr$ TEST $|I| - \delta$
INTERPOLATE NEW ϵ_{aa}
−
+
−
CALCULATE $\int (P^*)^2\,dr$ TEST $|I| - \delta'$
CALCULATE $P^* = P + \epsilon_{a\beta} S$
INTEGRATE (19) $S = S^+ + xQ$
+

Fig. 3. Block diagram of the steps involved in iterating one Hartree-Fock atomic-wave function

$$P_{i+1} = Z_{i+1}\left(1 + \frac{1}{12}\,h^2 g_i\right)^{-1} \tag{12A}$$

where

$$Z_{i+1} = (2 - G_i)Z_i - Z_{i-1} - F_i \tag{12B}$$

F and G are defined by the following equations:

$$G_i = h^2 g_i\left(1 + \frac{1}{12}\,h^2 g_i\right)^{-1} \tag{12C}$$

$$F_i = \frac{h^2}{12}\,(f_{i-1} + 10 f_i + f_{i+1}) \tag{12D}$$

The first neglected term of this difference equation has been shown to be (1/240) $\delta^6 P_i$.

In solving equation 1 in parts, a solution is first determined for the follow-equation:

$$\frac{d^2}{dr^2}P^+ + gP^+ + \phi = 0 \tag{13}$$

g and ϕ are defined in equations 5 and 8. This equation is made to fit the boundary conditions and the condition of normality in the following manner. The boundary condition at infinity is replaced by the condition

$$P(r_n) = a \tag{14}$$

where r_n is the value of r at the end of the last interval and is taken large enough so that a is a very small number. The integration is begun at the origin and an initial slope is assumed by choosing a value for P_1^+. In order to meet the condition expressed in equation 14, a complementary function is computed which is the solution of the corresponding homogeneous equation.

$$\frac{d^2}{dr^2}Q + gQ = 0 \tag{15}$$

The function

$$P = P^+ + xQ \tag{16}$$

is also a solution of equation 13. x may be chosen so that equation 14 is satisfied. It should be noted that a value of $\epsilon_{\alpha\alpha}$ has been assumed to calculate g. This solution is checked to see if the condition of normality is satisfied. If not, $\epsilon_{\alpha\alpha}$ is adjusted and equation 13 is integrated again. This cycle is repeated until the condition is satisfied.

At this point the solution must be made orthogonal to one or more other wave functions. In practice not more than one other function is involved to any appreciable extent. If P_β is the function to which P_α must be made orthogonal, then subtraction of equation 13 from equation 1 leaves the equation

$$\left(\frac{d^2}{dr^2} + g_\alpha\right)(P_\alpha - P) - \epsilon_{\alpha\beta}P_\beta = 0 \tag{17}$$

where P is the solution expressed in equation 16 and P_i is the solution required by equation 1. If S is defined as follows:

$$S = (P_\alpha - P)/\epsilon_{\alpha\beta} \tag{18}$$

Then equation 17 becomes

$$\frac{d^2}{dr^2}S + gS - P_\beta = 0 \tag{19}$$

which may be solved by the same program as was equation 13. Then the final solution of equation 1 is a linear combination of P and S.

$$P_\alpha = P - \epsilon_{\alpha\beta}S \tag{20}$$

$\epsilon_{\alpha\beta}$ is chosen so that P_α is orthogonal to P_β

$$\epsilon_{\alpha\beta} = \frac{\int_0^\infty P(r)P_\beta(r)\,dr}{\int_0^\infty S(r)P_\beta(r)\,dr} \tag{21}$$

Usually P_α is still very closely normalized. If this is not true, for a particular case, $\epsilon_{\alpha\alpha}$ must be readjusted at this point, and the equations integrated again.

Programming

The IBM type *650* is capable of storing 2,000 words on a magnetic drum memory. Sixty-four has been determined to be a sufficient number of values of r to yield an accurate solution with the difference equations which are used. With this many values of r it is possible to handle up to six shells of electrons without loading information into the machine in parts. Together with associated constants the input data require approximately 600 storage positions. Another 260 positions are needed to store intermediate results such as P^+ and g. Standard read and punch routines require about 100 positions. This leaves a little more than 1,000 words for the program. The present program uses almost all of this space.

A flow sheet is shown in Fig. 3 which indicates in block outline the steps taken to re-estimate the solution for one of the shells. The maximum value of any of the wave functions for any ion in the lighter end of the periodic table has well-defined limits which are separated considerably less than two orders of magnitude. For this reason, by careful planning of the sequence of the arithmetical steps, the whole program may be written in fixed point form.

At present the time required to complete one loop of the program requires from 15 to 25 minutes of machine time. By extrapolating from known solutions of other near-by ions it is usually possible to estimate starting functions closely enough (i.e., within a few per cent) to arrive at a self-consistent solution (within .01 or .02 per cent) after two or three iterations of each shell. The least tightly bound shell sometimes requires an additional one or two iterations. Thus the radial charge distribution of an ion with six shells of electrons may be obtained in approximately 5 hours of machine time. No attention has been given to optimizing the program by placing sequential orders that are used frequently in storage positions separated by an appropriate storage access time. This would shorten the running time severalfold.

The IBM *650* appears to be quite well suited to these self-consistent field calculations. At present it is necessary to watch the solutions as they progress through each iteration and exercise a certain amount of judgment as to the sequence of the iterations and the damping factor to be used in estimating the next trial solution. A faster machine would use an inordinate percentage of the time reading and writing this information. The IBM *650* uses 10 to 20 per cent of the time performing these input-output operations.

Conclusions

There is a need for readily available and accurate atomic-wave functions for theoretical studies of luminescent systems and other solid-state problems. The Hartree-Fock equations will produce good wave functions but are quite tedious to compute by hand methods. With a machine of intermediate speed and reasonably large fast-access memory, such as the IBM type *650*, it is possible to compute these wave functions in a few hours of machine time. A program to compute these functions is now operating so that any future needs for the wave functions of ions not yet computed can be made available in a short time.

References

1. THE WAVE MECHANICS OF AN ATOM WITH A NON-COULOMB CENTRAL FIELD, PART I. **D. R. Hartree.** *Proceedings*, Cambridge Philosophical Society, London, England, vol. 24, 1927, pp. 89–110.

2. ON THE NUMERICAL SOLUTION OF GENERALIZED EQUATIONS OF THE SELF-CONSISTENT FIELD, **V. Fock, M. J. Petrashen.** *Physikalische Zeitschrift der Sowjetunion*, U.S.S.R., vol. 6, 1934, pp. 362–414.

3. AN ABSOLUTE THEORY OF SOLID-STATE LUMINESCENCE, **Ferd E. Williams.** *Journal of Chemical Physics*, New York, N. Y., vol. 19, Apr. 1951, pp. 457–66.

4. NUMERICAL SOLUTION OF DIFFERENTIAL EQUATIONS (book), **W. E. Milne.** John Wiley & Sons, Inc., New York, N. Y., 1953, pp. 90–92.

Practical Applications of Electrostatic Phenomena to Particulate Matter

OLIVER C. RALSTON
NONMEMBER AIEE

THE behavior of particulate matter in electrostatic fields has been utilized in a number of industries; some are well known and others less known. A comprehensive picture cannot be put into one short paper, but several aspects of electrostatic technology deserve comparison and correlation. Those processes chosen for consideration in this paper are indicated in the headings.

Electrostatic Separation of Solid Particles

There are six or more electrostatic processes for separating mixtures of solid particles, but only the process using differences in conductance of the solids is widely known. All use an electrostatic field, commonly pulsating direct current, but sometimes alternating current, at high voltage, applied between two electrodes. In some instances the mixture falls in free air between the electrodes, and in some it travels into the field on a horizontal rotary cylinder or slides down a chute which acts as one of the electrodes, charged with respect to earth but usually grounded. Conductance separators utilize the difference in conductance of mineral particles, particles of high conductance acquiring a charge by contact with a charged conductor, or by being sprayed with a corona discharge. Frictional electric charging caused by rubbing of solid particles against each other, or against the feeder, or against the chute before reaching the electric separating field or while passing through it, is usable but less known. Differences in dielectric constant are utilized in other separators; see Fig. 1.

In Fig. 1, the field between electrodes is occupied by air or by a liquid of dielectric constant between two particles to be separated. Charged particles migrate toward oppositely charged electrodes, plus and minus. Dipoles migrate into the densest field. Particles of high dielectric constant compared to medium migrate toward the dense field; those of lower dielectric constant than the medium, away from the densest field.

Conductance electrostatic separators have been in use for nearly 50 years, and several firms sell machinery of this type. They operate by conveying a mixture of solid particles of markedly different conductance, such as metal sulfides (conductive) and silicates (poor conductors), into a zone where the charge can be fed to the conductive particles and simultaneously or shortly thereafter subjected to a static field of reasonably high intensity. Passage through the separating field takes place while the mixture of solid particles may be sliding down a chute or falling freely in air, or being conveyed over a charged or grounded roller, or carried on a belt. Charged particles of a definite sign are attracted toward the oppositely charged electrode that helps define the geometry of the field or is repelled by the electrode of the same sign of charge; see Fig. 2.

Fig. 2 (A) shows a model with a purely static field. Black particles are poor conductors and drop off the charged feeder electrode without deflection into the collecting bin *B*. Bin *A* is the feed bin. Particles of conductive minerals, denoted by hollow circles, take on the same charge as the carrier electrode and are repelled toward the other electrode to drop into bin *C*. In Fig. 2 (B) is shown a spray or corona discharge from a wire or knife edge pinning poor conductors to the carrier electrode.

Several dozen forms of apparatus merit attention. The means of getting electrostatic charge onto conductive solid particles is also quite variable. The mixture may rest on a conductive chute, roller, or belt that is charged with positive or negative electricity, or it may rest on a conductive support that is grounded but nevertheless has earth potential toward an insulated, highly charged electrode that forms the other boundary of the electric field. Only the good conductive particles receive effective charge in amounts to permit them to jump from their support, or be deflected from a falling trajectory by the presence of a highly charged electrode of opposite polarity. A corona or spray charge can

---

Paper 56-114, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 24, 1955; made available for printing December 7, 1955.

OLIVER C. RALSTON is with the U. S. Bureau of Mines, Washington, D. C.

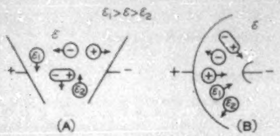


Fig. 1. Forces exerted in an electric field: horizontal section, looking down. Marked differences shown in field intensity of separators

be directed onto particles carried on conductive chutes, rollers, or belts; if the latter are grounded, the conductive particles at once lose their charge to ground, whereas the poor conductors cannot do this and they become electrostatically pinned to the carrier electrode.

If two different types of particles in a mixture differ markedly in dielectric constant they can be said to have different electron atmosphere pressures within them, and on rubbing together and then separating them it would be found that electrons have flowed from one type of particles to those of the other type. In common terms, frictional electricity has been created, with one set of particles positively charged and the other negatively charged. Dropping them through a static field will cause positive particles to deflect toward the negative electrode and negative particles toward the positive electrode, thus permitting a separation.

Particles that are essentially dipoles (charged oppositely at the two ends) before or while they are in a high-intensity field, tend to migrate from zones of low field intensity to those of high field intensity. This calls for apparatus wherein one electrode may be a plate of extensive area and the oppositely charged electrode may be one of very small area, such as a wire, around which field intensity is very high. The particles may be dropping vertically through such a field and the electrodes aligned in vertical position to allow deflection of the most pronounced dipoles to the zone of intense field.

Heating certain minerals, like quartz, causes pyroelectric polarization so each particle becomes effectively a dipole. Other associated minerals may not be affected. Semiconductive mineral particles entering an electric field permit flow of electrons to the side or end of the particle nearest the positive electrode, leaving the other side positively charged, and the particle has become a dipole by induction.

In practice, the coarsest mineral particles that can be separated reliably are of 20- to 10-mesh size, depending on their specific gravity. The smallest size that can be treated commercially is 100- to 200-mesh. A mass of mixed particles passing through an electrostatic field must be spread out in layers not more than one particle deep so that their individual trajectories cannot intersect. Smaller particles could be treated, but the limitation in area of the separating zone usually makes it uneconomical to treat them. Smaller sized dust tends to adhere to surfaces of apparatus or of larger particles, and therefore it is usually the practice to remove the dusts of finer size from the mixture to be treated. Designs of machines adapted for treating fine dust have been described, but they have not reached commercial adoption.

In Florida two very large commercial installations of conductance-type electrostatic separators with corona charging, work is being done in treating the family of heavy minerals removed from certain sea sands. The titanium minerals, ilmenite and rutile, happen to be the most conductive and are separated from zircon, monazite, quartz, and other minerals of lower conductance (and dielectric constant). Stainless-steel grindings are separated from dirt, sand, and abrasive grits of certain machine shops. Mustard seed is separated from twigs, leaves, sand, dead insects, insect and mouse excreta, and dried dead seeds. Chromite can be separated from the associated silicate minerals. In recent years, pairs of minerals that are not normally separable by any of these processes, if conditioned with a reactive gas, like hydrogen fluoride are surfacially altered to mixtures of good and poor conductors; an example is a mixture of quartz and feldspar. The possibilities of special surface conditioning to enhance differences in physical properties are almost endless and, in their versatility of application, they put electrostatic separations into the class of flotation processes.

Clarification of Aerosols, Cottrell Process

Gases carrying dusts, fogs, fumes, and smokes are common in industry and are frequently the source of economic losses or of nuisance. The process of clarifying these gases is known in the United States as the Cottrell process after the American inventor, and in England as the Lodge process, after the coincident British inventor. The object is to remove the dispersed solid or liquid phase from the true gas. The particles are charged from fine wires at high potential sufficient to cause a corona discharge, positive or negative. The gas and suspended matter become charged by the corona and then migrate at high speed to the oppositely charged electrode where they lose their charge during deposition or contact. The collecting electrode is of large area, and the deposited material is removed by dropping off, or by tapping the collecting electrode, or by keeping the latter flushed with a liquid, usually water, which washes them down.

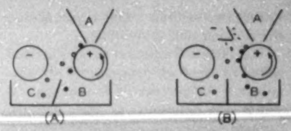


Fig. 2. Conductance separators. Vertical section through two electrodes made of rollers on horizontal axis and oppositely charged

A—Model with purely static field, showing poor-conductive material dropping off without deflection

B—Same as (A), showing poor-conductive material pinned to carrier electrode

The process is widely used and fairly commonly known, but most of the literature avoids discussion of principles and consists mainly of descriptions of installations, their operating details and their efficiency of operation. The forms of apparatus, Fig. 3, are the subject of an extremely extensive patent literature but they normally consist of vertical tubes through which the gas cloud passes, and their inner surfaces act as the grounded collecting electrode while an axial wire acts as the highly charged corona or charging electrode. This, now in wide use, was the earliest form of apparatus, known as the wire-in-pipe form of "treater." The wire-to-plate apparatus consists of vertical plates alternating with rows or curtains of vertical wires; it is less expensive to construct and can be installed in flues, or in expanded flues with the vertical plates parallel to the flow of gases. A third form sometimes used for cleaning the last traces of suspended matter consists of alternating curtains of charged corona wires and grounded collecting wires. Treaters, with flushed collecting electrodes are usually more efficient in collecting dusts of poorly conducting material, providing there is no objection to the collected material being caught in slurry form. "Electric precipitators" are usually more expensive than some of the mechanical gas-cleaning equipment, but freqeuntly they are so much more efficient as to be indispensable. They have had growing use over 40 years.

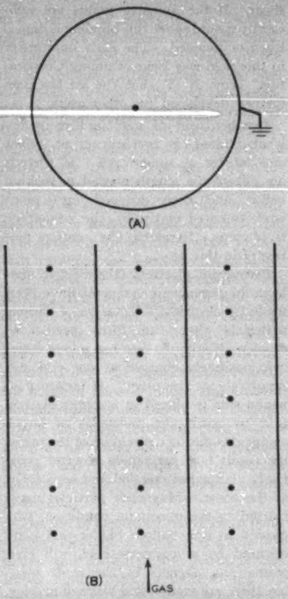


Fig. 3. Cottrell precipitators, horizontal sections. Pipe or plate grounded. Wires charged to corona potential, plus or minus. Power supply, rectified a-c sine wave

A—Field of wire-to-pipe design, vertical flow
B—Field of wire curtain-to-plates, horizontal flow

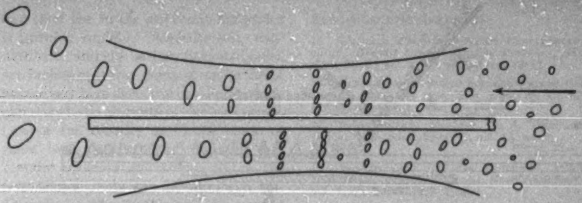


Fig. 4. Electric demulsifier, vertical section. Venturi tube electrode, axial rod electrode; a-c power supply. Water droplets in petroleum coalesced by "chaining" of ovoids of water

Breaking of Water-in-Oil Emulsions

Another invention of Cottrell, put to use during the same period as were gas clarifiers, is the separation of water from petroleum. Oil wells usually produce some water along with the crude petroleum; from various causes, including pumping, some of the water becomes finely divided, and a stable emulsion of water (often a brine) in oil results. Both chemical and electric coagulation processes are in wide use in the United States and around the world, and each has its particular advantages and field of usefulness.

The commonest form of treater for dehydrating such oils can be pictured, see Fig. 4, as a venturi tube connected to the ground, and an axial rod or wire electrode at high potential. Emulsion flowing along such a tube is subjected to a high-potential field; and the droplets of water, being the most conductive component, are distorted into long ovoids oriented toward the charged electrode. The distorted droplets tend to gather into chains between the electrodes separated by thin films of oil, and finally a spark flashing through the chain permits coalescence of the individual small droplets into a few much larger drops that are heavy enough to settle by gravity from the oil phase. The oil layer may be sufficiently treated by one pass or it may require retreatment by several passes through succeeding fields. Petroleum carrying salt or dirt can be churned with water to collect these impurities, and passed through the electric emulsion breaker; the impurities are thus washed out. Similar types of equipment have also been devised for separating precipitated wax from oils, but specialized designs are more efficient.

Either direct or alternating current can be used for this type of separation; high-tension alternating current, being less expensive, is almost invariably used. The Cottrell patents on this type of process were assigned to the California Petroleum Rectifying Corporation, which operated under that name but was ultimately merged with one of the companies using chemicals for coagulation of water droplets.

Electrostatic Spraying

When one of the electrodes creating a high-potential field has a sharp point or edge, at a certain potential drop a corona discharge appears near the sharp edge. When the potential is raised, sparks begin to displace the corona and, on further raise of potential, arcing begins to the opposite electrode. A liquid applied to the small, sharp electrode in the corona state of discharge is sprayed in droplets that are of the same electric sign as that of the corona electrode; the field guides the droplets to the opposite electrode. If the corona electrode is incorporated in a paint spray gun and the articles to be painted are connected to the opposite electrode, or interposed in front of it, the droplets of paint, all being charged and hence repelling each other, form a uniform cloud that precipitates (as in the Cottrell process) uniformly over the surface of the collecting electrode (the article to be painted). The spray gun may be provided with the corona electrode within the gun itself, or the corona electrode may be a grid of small wires close to the spraying nozzle, with the spray passing through the corona zone. Air or steam atomization may be used, or the nozzle may have an axial corona electrode; then all of the atomization would take place by electrical force. The combination of air atomization and electric migration and steering to the target is favored.

Processes of this type have come into wide use for spraying paint uniformly over surfaces with less loss due to overspray that misses the article. The electrical forces deflect much of the overspray to the article, and part of the overspray will even travel to its back side. Spraying of vitreous enamel slip on metal articles to be fired into hard-enameled parts of articles like refrigerators is becoming common. These spraying processes are best adapted to continuous production-line operations but are adapted to intermittent use as well. Spraying of oil onto a ribbon of tin plate passing into a tinning bath and many similar tasks are possible. The principal enterpriser in this field has been Ransberg Electro-Coating Corporation of Indianapolis, Ind.

Drainage of paint to the lower sharp edges of coated articles results in tears or drips of excess paint that are undesirable. If the article is an electrode in a field (either the grounded one or the one at high potential relative to earth), the concentration of field near its sharp edge causes reverse atomization of the tears. If corona potential is not reached, there is another electrostatic effect of lowering the apparent surface tension of a liquid in an intense field so that the tears are

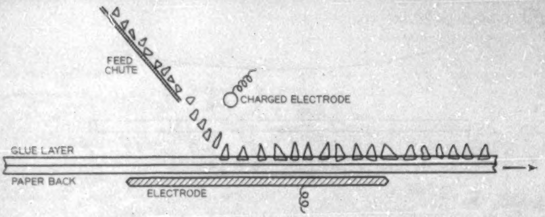


Fig. 5. Electrically oriented sandpaper. Vertical section of continuous machine; sharpest points of abrasive point to densest field

greatly decreased in size by their dripping off. Detearing by electrostatic means is also commercially used.[18]

Field strengths used in these processes vary from 30,000 to 100,000 volts per foot of spacing between electrodes. The field is usually pulsating direct current which is produced by rectification of high-potential alternating current from a variable voltage transformer.

Electrospinning of Fibers

Although commercial electrospinning of fibers has not as yet been attempted, as far as the writer can determine, serious research and development work have been conducted thereon. A list of United States patents in the selected bibliography shows that electrostatic forces were proposed as early as 1902 by Cooley as an aid in spinning artificial fibers. Norton was active in the United States and Formhals from 1934 to 1943 carried forward the development. Formhals took out at least eight United States patents that were ultimately vested in the Alien Property Custodian during World War II. Lastly, Childs assigned to Eastman Kodak Company, Rochester, N. Y., a patent issued in 1944 covering an invention that was an improvement over the others previously issued.[32]

Viscous liquid products, like cellulose acetate in acetone solution, or heat-hardening plastics, or molten glass, issues from a spinneret or other opening and is drawn toward charged electrodes by being charged oppositely in the stream of liquid or through having the spinneret charged. Electric attraction of the advancing fiber tends to stretch and attenuate the fiber before it has firmly frozen or set. Smooth fiber or crinkly shapes can be obtained by rapidly varying electric field or other factors during spinning. The series of patents makes instructive reading.

Electro-Orientation

A wide variety of processes for orienting particulate matter has come into use. Perhaps the earliest application was the preparation of oriented sandpaper. Abrasive grains are normally dropped onto glue on a paper or cloth backing and when the glue is hardened the result is sandpaper or emery cloth. Superior products were produced when the abrasive grains were fed into an electrostatic field maintained between the feeding device and the soft adhesive spread on the backing, or when an electrode was mounted just under it; see Fig. 5. If the most intense field is on the edge of the feeding device or near it and the adhesive-coated fabric is the opposite electrode (usually grounded) the grains of abrasive always have one point or edge that is sharper than any other area on the grain. Being at earth potential when fed in, if they actually contact the high-intensity electrode they take on charge by conduction. If they do not contact it, the sharp points on the particles encourage corona discharge from the charged electrode to the points, and the particles acquire charge of the same sign as the electrode and are repelled toward the coated fabric. The point of the particle is kept oriented toward the most intense field, while it drops, butt end down, into the adhesive film. After drying, the sandpaper thus has the sharpest points or edges facing away from the fabric and is ready to do the most efficient cutting of the article to be sandpapered. The product has another improvement over ordinary sandpaper, resulting from the fact that each abrasive particle is charged during its flight and the charges are all of the same sign so that they are repelled from each other and fall individually onto clean areas of adhesive, uniformly spaced from each other. Closely spaced grains on a sandpaper are often quickly blinded by accumulated cuttings being forced between them. If the abrasive grains are sufficiently spaced apart, this blinding is diminished or avoided. The early enterpriser in this field was Behr-Manning Corporation, Troy, N. Y., but the art has been further developed by other users.

Another field of manufacture, somewhat related, is preparation of velvet and fabrics of similar pile. If threads are cut of the length needed in making velvet and then dropped into place while oriented vertically, the advantage is obvious. Research Corporation pioneered in this.

Permanent electrets (the electric analogue of permanent magnets) have been made by electrical orientation of various materials that can form permanent dipoles. Carnauba wax has a long molecule positively charged at one end and negatively at the other. If melted carnauba wax is placed in a strong electric field to orient the molecules or larger aggregates during congealing of the mass, the result is a permanent electret positively charged at one end and negatively at the other. Magnetite particles magnetized to the maximum condition, suspended in any molten binder and kept oriented by a magnetic field, will give permanent magnets in the same way, but there are many other ways of making them. The field for permanent magnets is growing, but there has been virtually no known way of making permanent electrets for a similar wide series of industrial applications.

The same means of orienting anisotropic particles in an electrostatic field during setting of a resin in a molten binder permits preparation of light polarizing articles or oriented sheets of condenser material like rutile particles (whose dielectric constant parallel to their minor crystallographic axis is 89 but is 173 when parallel to the major axis). This latter article leads to lowest dielectric loss and reduction in power factor, according to the Western Electric Products, New York, N. Y.

Conductivities of gold sols increase in an electrostatic field, and it has been shown that particles have oriented into countless chains of metal particles, which accounts for increased conductance.

The electroviscous effects in liquids and suspensions of solids in liquids are caused by a number of factors whose effects are additive. Depending on the geometry of flow an increase or decrease in viscosity can take place due to electro-orientation. Flat-flake and needle shapes contribute to this effect.

The patent literature abounds in examples of applications of such phenom-

Vol. 7
Pt. 1
1956

ena. Jacobs is interested in hardenable plastics reinforced by flake mica. If, during heating of the plastic in the liquid condition, the mica flakes are electrically oriented, the final hard product has a dielectric strength normal to the faces of the oriented flakes which is different from that parallel to them. A particularly clever application of electrostatics is that by Gwyn and Williams for packing sheet glass for shipment with sawdust or similar soft material for filler between the sheets of glass to prevent rubbing; see selected bibliography. The sheets of glass approaching the packer's box on a conveyor belt are charged by contact with an electrode at say 70,000 volts or rubbed by a silk roller to charge the surface. Sawdust sprinkled over the surface adheres until the sheet of glass is placed and another similarly placed on top of it. Dissipation of the charge leaves the glass packed with a minimum of spacing between sheets and yet adequately protected against rubbing together during transit or handling.

The use of electrified particles to point out the location of flaws in enamels and glazes was indicated by Jones, see selected bibliography, and developed for commercial use by the Magnaflux Corporation of Chicago, Ill., with the trade name of Stratiflux. Steel, coated with vitreous enamel, can be positively charged and fine talc dust suspended in air blown past a negative electrode onto the enamel surface. The electrified dust adheres only to cracks or other flaws in the enamel. If the body coated with enamel is nonconducting, a special penetrating aqueous solution is absorbed through the defects first.

All of the applications of electroorientation cannot be covered here; however zerography or "electrical printing and copying" should be mentioned as developed mainly at the Battelle Institute, Columbus, Ohio, which makes a more complex application of the phenomena just mentioned.

Piezoelectric solids become charged into dipoles, oppositely charged at opposite ends by being stressed. Pyroelectric phenomena are closely related and probably depend on thermal stresses bringing about piezoelectric charging. The reverse phenomenon is electrostriction or changes in dimension of aniosotropic crystals with different dielectric constants in the different crystallographic directions when an electric unidirectional field is impressed on the crystal. The ferroelectric semiconductors of recent development are materials of this kind and their use in the electronic industries is growing rapidly. Rochelle salt, seignett salts, and the barium titanate bodies are included in this group of materials of excessively high dielectric constants which can store large amounts of electric energy. One of the theories as to the cause of these high dielectric constants is that certain electrons in the atoms of the ferroelectric body can be displaced sufficiently by the electric field to convert the body to a mass of tiny atomic dipoles all oriented similarly until the imposed electrostatic field is removed.

What has been said gives point to the thought that electrostatics, a relatively neglected field in electrical technology, has been put to increasing use by industry and that the field of further application and invention is still quite uncultivated.

Selected Bibliography

ELECTROSTATIC SEPARATION OF PARTICULATE SOLIDS

1. The Scope and Economics of Electrostatic Separation, **L. H. Bullock.** *Industrial and Engineering Chemistry*, American Chemical Society, Easton, Pa., vol. 33, 1941, pp. 1119–23.

2. Grinding Dust Separation, **A. L. Field.** *U. S. Patent no. 2,388,731*, 1941.

3. Dielectric Constant in Air-Ambient Electrostatic Separation, **F. Fraas, O. C. Ralston.** *Report of Investigations 4278*, U. S. Bureau of Mines, Washington, D. C., 1948, 12 pp.

4. Universal Type Electrostatic Separator, *Ibid., 4766*, 1951, 4 pp.

5. Electrostatic Separations of Solids, *Ibid. Industrial and Engineering Chemistry*, American Chemical Society, Easton, Pa., vol. 32, 1950, pp. 600–04.

6. Conductance Electrostatic Separation with Convective Charging, *Ibid., Transactions*, American Institute of Mining and Metallurgical Engineers, New York, N. Y., vol. 187, 1950, pp. 1041–43.

7. Lamellar Grading of Powdered Mica, *Ibid. Report of Investigations 3662*, U. S. Bureau of Mines, Washington, D. C., 1942, 7 pp.

8. The Electrostatic Separation of Several Industrial Minerals, *Ibid., Technical Publication 2408*, American Institute of Mining and Metallurgical Engineers, New York, N. Y., 1948, 11 pp.

9. Contact Potential in Electrostatic Separation, *Ibid., Report of Investigations 3667*, U. S. Bureau of Mines, Washington, D. C., 1942, 17 pp.

10. An Electrostatic Separator for Fine Powders, *Ibid., 3677*, 1942, 6 pp.

11. Electrostatic Separation, Parts I, II and III, **H. B. Johnson.** *Engineering and Mining Journal*, New York, N. Y., vol. 139, Sept. 1938, pp. 42–45; Oct. 1938, pp. 42–43; Dec. 1938, pp. 41–45.

12. Notes on Electrostatic Separation of Diamonds, **A. A. Linari-Lindholm.** *Journal*, Chemical, Metallurgical and Mining Society of South Africa, Johannesburg, Union of South Africa, vol. 52, Aug. 1952, pp. 33–35. Also *Industrial Diamond Review*, Henry Paulson and Company, Chicago, Ill., vol. 11, Apr. 1951, pp. 89–92; May 1951, pp. 110–22.

13. The Electrostatic Preparation of Brown Coals, **E. von Szantho.** *Das Braunkohlenarchiv*, Halle, Germany, No. 56, 1949, pp. 29–48.

14. Electrostatic Beneficiation of Stones and Earth, *Ibid., Tonindustrie-Zeitung und Keramische Rundschau*, Wilhelmshaven, Germany, vol. 77, 1953, pp. 83–88.

ELECTROSTATIC SPRAYING

15. Wrinkle Finishes Applied Electrostatically, **Herbert Chase.** *Organic Finishing*, New York, N. Y., vol. 9, No. 1, 1948, pp. 33–37.

16. Electrostatic Spray Method Primes Car Bodies, *Ibid., Iron Age*, New York, N. Y., vol. 172, no. 14, Oct. 1, 1953, pp. 102–03.

17. Electrostatic Glazing, **T. A. Dickinson.** *Ceramic Age*, Newark, N. J., vol. 54, no. 1, 1949, pp. 28–29.

18. Electrostatic Coating Processes, **Arthur P. Schulze.** *Organic Finishing*, New York, N. Y., vol. 7, no. 6, 1946, pp. 9–22.

19. Electrostatic Atomization Applied to Paint Spraying. *Steel*, Cleveland, Ohio, vol. 129. Aug. 6, 1951, p. 80.

ELECTROSPINNING OF SYNTHETIC FIBERS

20. **J. F. Cooley.** *U.S. Patent no. 692,631*, 1902.

21. *Ibid., U.S. Patent no. 749,276*, 1903.

22. **A. Formhals.** *U.S. Patent no. 1,975,504*, 1934.

23. **C. L. Norton.** *U.S. Patent no. 2,048,651*, 1936.

24. **A. Formhals.** *U.S. Patent no. 2,077,373*,1937.

25. *Ibid., U.S. Patent no. 2,109,333*, 1938.

26. *Ibid., U.S. Patent no. 2,116,942*, 1938.

27. *Ibid., U.S. Patent no. 2,123,992*, 1938

28. *Ibid., U.S. Patent no., 2,158,415* 1939.

29. *Ibid., U.S. Patent no. 2,158,416*, 1939.

30. **C. L. Norton.** *U.S. Patent no. 2,185,417* 1940.

31. **A. Formhals.** *U.S. Patent no. 2,323,025*, 1943.

32. **H. R. Childs.** *U.S. Patent no. 2,338,570*, 1944.

ELECTRO-ORIENTATION

33. Sand Paper Grows Up, **D. E. Killefer.** *Industrial and Engineering Chemistry*, American Chemical Society, Easton, Pa., vol. 29, 1937, pp. 849–54.

34. Electrocoating. A Method of Deposition, Orienting and Dispersing Finely Divided Material by Means of Electrostatic Field Forces, **J. O. Amstutz.** *Rubber Age*, New York. N. Y., vol. 49, 1941, pp. 19–22.

35. Electrocoated Pile Fabrics, **N. E. Oglesby, L. E. Hoogstoel.** *Industrial and Engineering Chemistry*, American Chemical Society, Easton, Pa., vol. 32, 1940, pp. 1552–54.

36. Use of Electrets in Electrical Instruments, **A. Gemant.** *Review of Scientific Instruments*, American Institute of Physics, New York, N. Y., vol. 11, 1940, pp. 65–71.

37. Electrets, *Ibid., Philosophical Magazine* London, England, vol. 20, 1935, pp. 929–52.

38. Permanent Charges in Solid Dielectrics. Part I, **B. Gross, L. F. Denard.** *Physical Review*, American Institute of Physics, New York, N. Y., vol. 67, 1945, pp. 253–59.

39. Part II, *Ibid., Journal of Chemical Physics*. American Institute of Physics, New York, N. Y., vol. 17, 1949, pp. 866–72.

40. Deposition of Material, **H. A. Wintermute.** *U.S. Patent no. 2,221,338*, 1940.

41. Apparatus for Coating Articles by Electrostatic Deposition, **A. Thode.** *British Patent no. 497,540*, 1938.

42. Condenser Material and Method of Making Same, **H. F. Fruth.** *U.S. Patent no. 2,384,541*, 1945.

43. Method of Making Insulating Material, **E. J. Jacob.** *U.S. Patent no. 2,430,123*, 1947.

44. Protection of Surfaces, **J. D. Gwyn, C. B. Williams.** *U.S. Patent no. 2,476,145*, 1949.

45. Electrostatically Charged Particles in the Detection of Minor Surface Defects in Porcelain Enamel, **W. H. Jones.** *Journal*, Canadian Ceramic Society, Toronto, Ontario, Canada, vol. 18, 1949, p. 29.

46. Electrified Particle Inspection of Ceramic Materials, *The Enamelist*, Cleveland, Ohio, vol. 27, nos. 2 and 3, 1950, pp. 37–38.

Transcendental Function Analogue Computation with Magnetic Cores

D. H. SCHAEFER
ASSOCIATE MEMBER AIEE

R. L. VAN ALLEN
ASSOCIATE MEMBER AIEE

SQUARE loop magnetic cores used in conjunction with switching transistors can be made to perform a great number of arithmetical operations. Previous papers have shown how these elements can be used to obtain 4-quadrant multiplication[1] and division[2]. This paper presents a basic circuit and shows how this circuitry can be utilized to obtain various sorts of computations. Specifically, experimental data are presented showing results obtained with the use of the circuit to obtain an output voltage proportional to the square root and other fractional powers of an input voltage, the arc sine of an input voltage, the sine of an input voltage and the product of two voltages. These are only a few of the possible outputs which are theoretically attainable.

The computing element described here is completely static in its operation. The output is in the form of current or voltage pulses whose average values are a specific function of one or two input voltages.

Basic Computing Circuit

Fig. 1 shows the basic circuit used for all the functions described in this paper. The magnetic core shown is of the "rectangular" or high-remanance type, aud type *p-n-p* transistors are used only as switches, appreciably enhancing their reliability and power-handling capacity. The voltage e_{ac} is a periodic voltage whose wave shape is dependent upon the computation desired and whose magnitude is great enough to drive the core from negative to positive saturation in a half-cycle. The voltage e_1 is in all cases a direct voltage of the indicated polarity. The voltage e_2 is usually a direct voltage, although for some computations it assumes other forms. When both of the input voltages e_1 and e_2 are direct voltages, the output voltage e_0 consists of a series of rectangular pulses whose height is proportional to e_2 and whose width is proportional to some function of e_1. Therefore the average output is proportional to $e_2 \times f(e_1)$, the particular function f depending upon the shape of e_{ac}.

In brief, the circuit operates as follows: When e_{ac} has the polarity shown, the emitter-collector resistance of transistor 1 is low, so the voltage e_1 is applied directly across the core windings and changes the flux level of the core from positive saturation to some flux level between positive and negative saturation, the amount of change being proportional to the magnitude of e_1. During this half-cycle, the emitter-collector resistance of transistor 2 is high and no voltage appears across the output resistor.

When e_{ac} has a polarity opposite to that which is shown in Fig. 1, the voltage e_{ac} itself is applied directly across the core windings in such a direction as to change the flux level of the core back to positive saturation. When it is remembered that e_{ac} has a value such as to enable it to reset completely the core on any given half-cycle, it is obvious that when e_{ac} is negative, the core will become saturated at some point during the half-cycle unless e_1 on the previous half-cycle has set the flux at negative saturation. During the portion of this negative half-cycle of e_{ac} that the core is unsaturated, the transformed voltage e_{ac} is applied to the base of transistor 2, allowing it to conduct, and therefore the voltage e_2 appears across the output resistor. After saturation of the core, only negligible voltage can be transformed to the base of transistor 2, therefore its emitter-collector impedance is high and the output voltage again falls to zero. The output voltage exists only when e_{ac} is negative and the core is unsaturated. How the shape of e_{ac} affects the length of time it will take to saturate the core during the negative half-cycle of e_{ac} will be examined in a moment.

To be convinced that the circuit will operate as just described, the reader should be aware of the following rule governing the switching characteristics of *p-n-p* transistors: If the base is more positive than both the emitter and collector, the emitter-collector impedance is very high, approximating an open circuit; if on the other hand, the base is more negative than both the emitter and collector, the emitter-collector impedance is near that of a short circuit as the transistor is saturated. If this rule is applied and if the effect of transformed voltages on transistor number 2 is kept in mind, it will be seen that the three modes of operation of the circuit are as shown in Figs. 2(A), (B), and (C), Fig. 2(A) being the equivalent of Fig. 1 when e_{ac} is positive; Fig. 2(B) the equivalent when e_{ac} is negative and the core is unsaturated; and Fig. 2(C) the equivalent when e_{ac} is negative and the core is saturated.

Theory of Computation

The deferred question of the time required to saturate the core on the negative half-cycles of e_{ac} will now be examined. It should be recalled that during the positive half-cycle of e_{ac}, the flux level of the core was changed by an amount $\Delta\phi$ that was proportional to the value of e_1 during that half-cycle. Therefore, during the negative half-cycle of e_{ac}, $\Delta\phi$ volt seconds must be "read out" of the core by e_{ac} before the core will saturate. This can be mathematically stated as

$$\int_0^{t_s} e_{ac}dt=\Delta\phi=ke_1 \qquad (1)$$

where time is measured beginning with the negative half-cycle of e_{ac} and t_s is the time at which the core saturates. It is wished to solve for the value of t_s. If, for a first example, e_{ac} is a square wave, i.e., if e_{ac} equals a constant K during each half-cycle, then equation 1 can be written as

$$\int_0^{t_s} Kdt=ke_1 \qquad (2)$$

$$K\int_0^{t_s} dt=ke_1 \qquad (3)$$

$$\text{or } t_s=\frac{ke_1}{K} \qquad (4)$$

Equation 4 states that the time necessary to saturate the core will be proportional to e_1. Therefore the width of the output pulse will be proportional to e_1. The height of the output pulse is e_2. The average value of the output voltage will therefore be proportional to the product of e_1 and e_2, if e_{ac} is a square wave. The circuit in this case is a single-quadrant version of the multiplier presented in reference 1.

Paper 56-134, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 21, 1955; made available for printing November 30, 1955.

D. H. SCHAEFER and R. L. VAN ALLEN are with the U. S. Naval Research Laboratory, Washington, D. C.

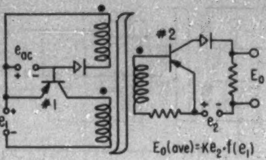


Fig. 1. Basic computer circuit

As a second example, let e_{ac} have the form Kt during the read-out half-cycle of e_{ac}. A triangular wave form will, for instance, provide such a ramp function. Equation 1 now becomes

$$\int_0^{t_s} Kt\,dt = ke_1 \tag{5}$$

$$\int_0^{t_s} t\,dt = \frac{ke_1}{K} \tag{6}$$

$$t_s^2 = \frac{2ke_1}{K} \tag{7}$$

$$t_s = \sqrt{\frac{2ke_1}{K}} \tag{8}$$

Under these circumstances, equation 8 shows that the time necessary to saturate the core will be proportional to the square root of e_1. Therefore the width of the output pulse will be proportional to the square root of e_1, and the average output will be proportional to $e_2 \times \sqrt{e_1}$.

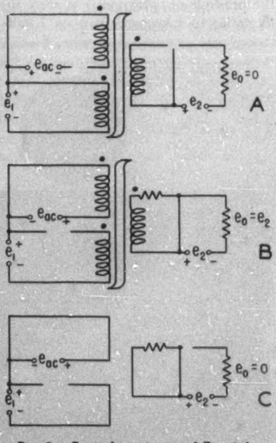


Fig. 2. Equivalent circuits of Fig. 1 for

A—e_{ac} positive
B—e_{ac} negative and core unsaturated
C—e_{ac} negative and core saturated

Table I. Output Voltages Produced by Various Wave Forms of e_2 and e_{ac}

e_1	e_2	e_{ac}	E_0 (ave)
d-c	d-c	square wave	$e_1 \times e_2$
d-c	d-c	Kt	$e_2\sqrt{e_1}$
d-c	d-c	Kt^n	$e_2 \times e_1^{\frac{1}{n+1}}$
d-c	d-c	$\cos \omega t$	$e_2 \times \sin^{-1} ke_1$
d-c	$\cos \omega t$	square wave	$\sin ke_1$
d-c	Kt^n	square wave	e_1^{n+1}
d-c	$f(t)$	square wave	$\int_0^{ce_1} f(t)dt$
d-c	Kt^n	Kt^m	$e_1^{\frac{n+1}{m+1}}$
d-c	$\cos \omega t$	Kt	$\sin k\sqrt{e_1}$
d-c	Kt	$\cos \omega t$	$(\sin^{-1} ke_1)^2$

Generalizing this example, it can be seen that if e_{ac} has the form Kt^n, the output will be proportional to e_1^x where the exponent x equals $1/(n+1)$. Letting n equal 2, for instance, will make the output proportional to the cube root of e_1. It will be shown that, experimentally, exponential outputs between 0.5 and 0.8 could be obtained with the use of e_{ac} wave shapes formed from resistance-capacitance networks modifying a square wave.

If one wishes to obtain an output proportional to the arc sine of the voltage e_1, it is necessary that e_{ac} be a cosine. If $e_{ac} = K \cos \omega t$, equation 1 becomes

$$K\int_0^{t_s} \cos \omega t\; dt = ke_1 \tag{9}$$

$$\frac{K}{\omega}\sin \omega t_s = ke_1 \tag{10}$$

$$\omega t_s = \sin^{-1}\frac{\omega k e_1}{K} \tag{11}$$

$$\text{or } t_s = \frac{1}{\omega}\sin^{-1}\frac{\omega k e_1}{K} \tag{12}$$

In practice, the cosine function can be easily obtained from a sinusoid voltage in series with a core that saturates in the middle of each half-cycle. This will be further detailed in the section on the practical circuit.

From this it can be seen that theoretically the output can be made proportional to any function of e_1 if e_{ac} can be formed into the shape of the derivative of the inverse of the desired function. For instance, if the voltage e_{ac} had the form $1/\sqrt{a^2 - t^2}$ the output would be proportional to the sine of e_1. To obtain such a time function is, to put the matter mildly, not easy. There is a relatively simple way, however, to obtain an output proportional to the sine of the input, and that is to let e_{ac} be a square wave and have the voltage e_2 assume the form of a cosinusoid. Having e_2 able to assume various wave forms in fact increases the number of functions easily obtainable by a large number.

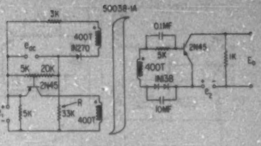


Fig. 3. Basic experimental circuit; core material Orthonol

The theory of the sine output element using a square wave e_{ac} and cosinusoid e_2 is as follows. Equation 4 states that with e_{ac} a square wave t_s will be proportional to e_1. The average voltage across the output resistor will be proportional to the integral of e_2 from zero to t_s. Mathematically

$$E_0(\text{ave}) = \int_0^{t_s} e_2\,dt \tag{13}$$

In the circuit under discussion, e_2 equals $C \cos \omega t$ and, from equation 4, $t_s = ce_1$ where C and c are constants. Therefore

$$E_0 = C\int_0^{ce_1} \cos \omega t\; dt \tag{14}$$

$$\text{or } E_0 = \frac{C}{\omega}\sin \omega c e_1 \tag{15}$$

Again the cosine function may be obtained by letting a sinusoidial voltage saturate a core.

If e_{ac} is a square wave and e_2 has the form Ct^n a derivation analogous to the foregoing shows that the output will be proportional to e_1^{n+1}. If e_2 is a ramp function, the output will be proportional to the square of the input, while if e_2 has a

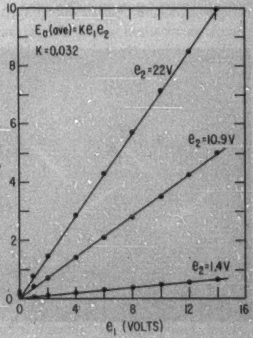


Fig. 4. Output proportional to product

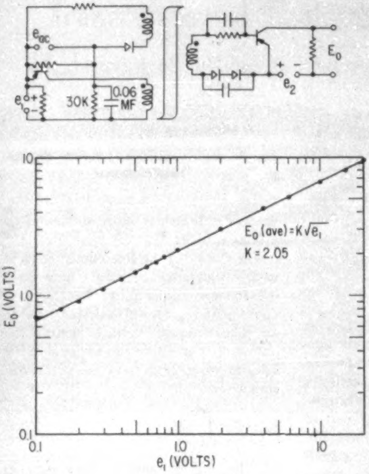


Fig. 5 (left). Output proportional to square root

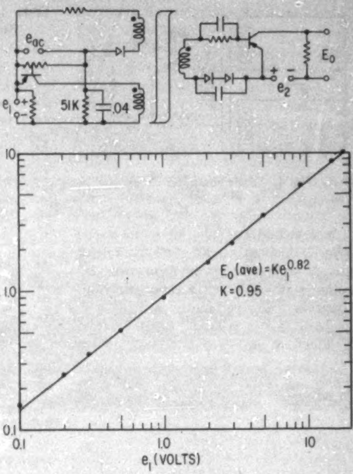


Fig. 7 (right). Output proportional to input raised to 0.82 power

parabolic wave form, for instance, the output will be proportional to the cube of the input.

Table I summarizes various combinations of voltage wave forms and the outputs they produce. Complex functions can be solved in two ways. One is to put functional wave forms on both e_{ac} and e_2. For instance, if e_{ac} has a trianglar wave form and e_2 a cosinusoidal form, the output will be proportional to the sine of the square root of e_1. If the two voltages are interchanged so that e_{ac} is cosinusoidal and e_2 is triangular, the output will be proportional to the square of the arc sine of e_1. As shown in the Table I, a wide variety of exponential outputs can be obtained by putting exponential wave forms on both e_{ac} and e_2.

The second method of building complex functions is to cascade elements. This can be done as it is the average value of e_1 that determines the amount of flux reset. By putting elements in series and by using each element to its full potentiality, exceptionally complex functions can be quickly synthesized. It should, of course, be realized that every added element in series delays the time in obtaining the answers by 1 half-cycle of the supply frequency.

Practical Considerations

The precision or accuracy of any of the computations performed by the analogue devices described in this paper are of the order of 1 per cent (%) or 2% of full-scale output. To insure this accuracy over the entire operating range, a reasonable amount of care should be taken with the choice of the various supplies; in particular, the source of e_{ac} and e_1. Since the principle of operation is to store information in a magnetic core on 1 half-

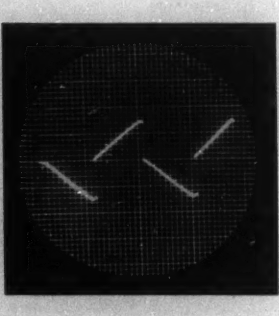


Fig. 6. Wave form of e_{ac} used for extracting square root

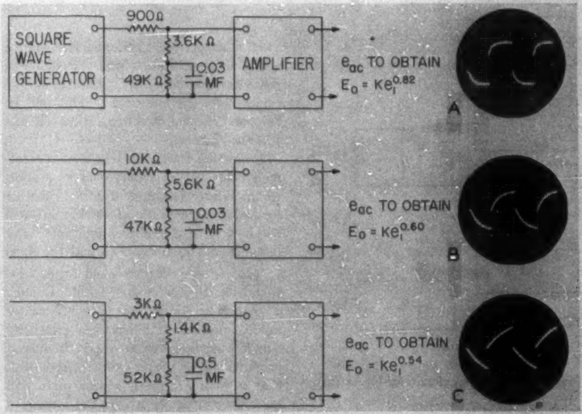


Fig. 8. Wave forms of e_{ac} and circuitry used to obtain this voltage for fractional exponents

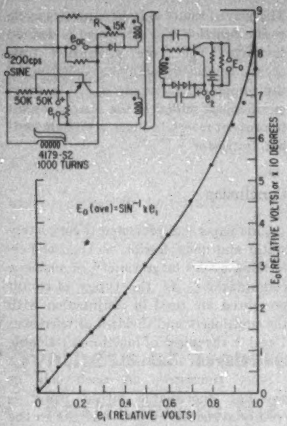


Fig. 9. Output proportional to arc sine

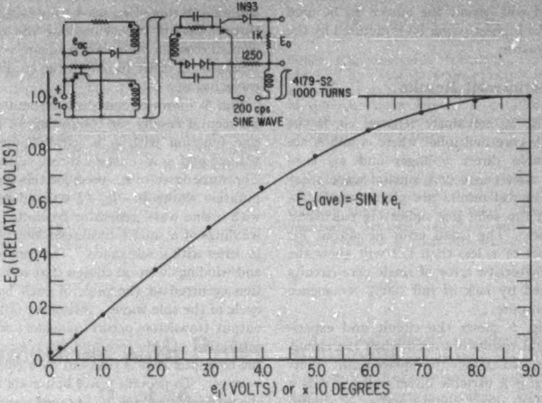


Fig. 11. Output proportional to sine

cycle and then to read out of the core the same information as a different function of time on the next half-cycle, a certain amount of work must be done in either case. Therefore the source voltage for read-in or read-out must be capable of delivering a small amount of power with good regulation. In the several experimental circuits tested, the input impedances have not been greater than several thousand ohms, and hence, the source impedance of the supplies have been usually 1,000 ohms or less.

When a voltage is applied to the windings of a square hysteresis loop core material, magnetizing current must be supplied before any flux change can be produced. This may have detrimental effects when the voltage e_1 is low and is supplied from a high-impedance source. The effect will be to delay the action of e_1 such that no output will be obtained until e_1 has been increased sufficiently to produce magnetizing current. Therefore it is necessary by some means to provide a magnetizing current bias. The resistor R in Fig. 3 has been added to the circuit primarily for this purpose; it is the actual circuit used for the multiplier and, with minor changes, for the other computations.

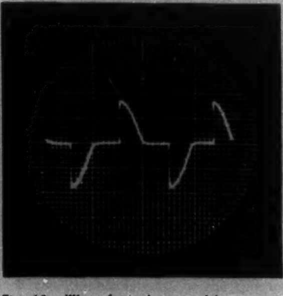


Fig. 10. Wave form of e_{ac} used for arc sine calculation

The effect of temperature on transistors used as switches is negligible, provided there is adequate switching voltages to hold the transistors in one of two states, saturation or cutoff. At high temperatures, the collector-cutoff current will increase appreciably unless a blocking voltage is provided which holds the base slightly positive with respect to the emitter. The switching transistor in the output circuit of the basic computer, Fig. 1, has no provision for a positive blocking voltage on the base. For reliable operation at elevated temperatures, it is necessary to add a small positive voltage of less than 1 volt which may be provided from a bias source; however, this requires an additional d-c source which can be eliminated by diodes in parallel with a capacitor, as shown in the output circuit of Fig. 3. The current pulses out of the base, switching the transistor charge the 10-microfarad capacitor to approximately 1 volt, which is twice the forward voltage drop of a single *1N138* silicon diode.

The 5,000-ohm by-pass resistor also shown in the base circuit provides base current limiting to the low-frequency components of the base voltage pulses transformed by the core. The 0.1-microfarad capacitor passes the high-frequency components to aid in precise time switching of the transistor.

The experimental computing circuits discussed in this paper have been operated at a supply frequency of 200 cycles per second. Difficulties arise when the frequency is increased in this single-core circuit, whereas, the 2-core, 4-quadrant multiplier[1] has been operated at 5 kc, maintaining an accuracy of ±1%. As explained in reference 1, the single-core multiplier has a small output for zero input at e_1 which is produced by the core having less than 100% remanence. This error appears in the output as narrow pulses, which at low frequency contribute less than 1% error but at higher frequencies may become an appreciable percentage of full output, unless some provision is made such as in the 2-core circuit to eliminate this difficulty. An operating frequency of 200 cycles per second was chosen to obtain these data in order to present the circuit in its simplest form and still exhibit 1% accuracy.

The experimental data presented in this paper were obtained at a maximum power level of 0.1 watt. The *2N45 p-n-p* type of transistor can be used to switch power to about 2 watts, provided sufficient base drive is available. This is an important factor to consider in proposed applications of this type of computer circuitry since in many control applications it is required to provide a unique control function at high-power levels. The proper choice of output transistor and power source from which e_2 is supplied will permit the output to be used at the highest power level dictated by the transistor.

Experimental Results

Fig. 3, previously referred to, is the single-core multiplier where e_1 and e_2 are variable direct voltages and e_{ac} is a 200-cycle-per-second square wave. Experimental results are plotted in Fig. 4 with the solid line indicating calculated values. The small error in output for values of e_1 less than 1/2 volt show the characteristic error of single-core circuits caused by lack of full 100% remanence of the core.

Fig. 5 shows the circuit and experimental results obtained when the circuit is used to extract the square root. Voltage e_1 is a variable direct voltage, e_2 is a constant direct voltage and e_{ac} is a ramp function obtained by adding a triangular and square wave form 90 degrees out of phase. The wave form of e_{ac} is shown in shown in Fig. 6. As previously mentioned, the wave form of e_{ac} could be a triangular wave; however, the wave form shown in Fig. 6 is a more efficient one for this application since the ramp function is continuous over a complete half-cycle. Actual observed points are plotted with the solid line indicating calculated values.

Fig. 7 shows the circuit and experimental results for a fractional exponent of 0.82; e_1 was a variable direct voltage with e_2 a constant direct voltage. Outputs proportional to input voltage e_1 raised to the 0.6 and 0.54 power were also experimentally obtained with different wave forms for voltage e_{ac}. Fig. 8 shows the wave forms of e_{ac} used to obtain the three different exponents and the networks necessary to modify a square-wave generator in order to obtain the required wave forms.

Fig. 9 shows circuit diagram and experimental results for computing an arc sine function with e_1 a variable direct voltage and e_2 a constant direct voltage. The wave form of e_{ac} used for this computation shown in Fig. 10 was obtained with a sine wave generator connected to windings of a square hysteresis loop core in series with a resistance. The core size and windings were so chosen that saturation occurred at the peak of each half-cycle of the sine wave. Switching of the output transistors occurs coincident with saturation of this core and for 1/4 cycle the computer core is read-out by a cosine voltage. To provide a full half-cycle for the read-in period for e_1, the input transistor is switched by the sine wave generator instead of the voltage e_{ac}. The compensating resistance R is shunted across the rectifier in this case for best results. Because the core does not possess 100% remanence and because of the nature of the arc sine function, a constant was added to the output which displaced the response curve in a positive direction equivalent to a phase shift of 5 degrees. To correct for this constant, a direct current was introduced in the output load resistance equal and opposite to the amount of the displacement.

Fig. 11 shows the circuit diagram and results of a computing circuit whose output is proportional to the sine of the input voltage e_1, where e_1 is a variable direct voltage. Supply voltage e_{ac} is a square wave as used for the multiplier. The wave form of e_2 is the same 1/4-cycle cosine function of Fig. 10 and is derived similarly. The core saturation time occurs coincident with the switching of the square-wave supply e_{ac}. An extra rectifier was added in the output circuit to prevent reverse voltage across the output transistor.

Conclusion

This paper has presented a completely static and quite flexible method of performing a very large variety of analogue computations. If the types of circuit presented are used in conjunction with the multipliers and dividers of references 1 and 2, the range of functions attainable is even further increased.

Since transistors are used only as switches, appreciable output power is readily available. Power outputs on the order of watts are obtainable, for instance, with nominal 150-milliwatt transistors.

In summary, a rugged, fast static method of performing analogue computation has been presented that is dependent only upon the wave form of periodic functions and the ability of a high-remanence magnetic core to store a given number of volt-seconds for a half-cycle of a supply frequency.

References

1. Four-Quadrant Multiplication with Transistors and Magnetic Cores, **R. L. Van Allen.** *AIEE Transactions*, vol. 74, pt. I, Nov. 1955, pp. 643–48.

2. A Static Magnetic Transistor Analogue Divider, **D. H. Schaefer.** *Ibid.*, pp. 652–55.

◆

Discussion

A. J. Schiewe (Purdue University, W. Lafayette, Ind.) and **K. Chen** (Westinghouse Electric Corporation, East Pittsburgh, Pa.): The authors are to be commended for their success in extending the pulse modulation technique to obtain transcendental function analogue computation with magnetic cores and switching transistors. The flexibility of the basic computer circuit is particularly impressive in producing a great variety of transcendental functions as listed in Table I. We have developed a number of analogue circuits different from the authors' to perform various transcendental functions, but also based on the principle of width and amplitude modulation of voltage pulses of specific shapes. An important difference between our circuitry and that of the authors, is that only switching transistors are used in our basic circuits. No magnetic cores are used because our experience with pulse-width modulation by means of magnetic cores made of presently available materials showed a lack of sharpness in the edges of the pulses, especially at frequencies around and above 1 kc. Since magnetic cores are not used in our basic circuits, no integration of voltage with respect to time is accomplished as given by equation 1 of this paper. However, transcendental functions can easily be generated with switching transistors alone. The wave forms shown in Figs. 12 and 13 illustrate the basic principles of our logarithmic and antilogarithmic circuits, respectively. In the logarithmic circuit, the direct input voltage E_i is superimposed on a train of exponential voltage pulses e_L. The difference between these voltages $(e_L - E_i)$, is used to control a

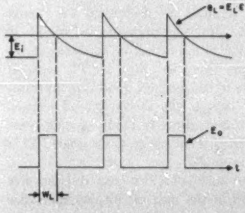


Fig. 12. Basic principles of the logarithmic circuit

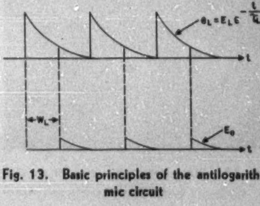


Fig. 13. Basic principles of the antilogarithmic circuit

transistor switch such that a constant voltage is switched on the load when (e_L-E_t) is positive and switched off the load when (e_L-E_t) is negative. The output voltage is, therefore, a train of rectangular pulses with fixed amplitude and frequency and with a width W_L given by the equation:

$$W_L=-\tau_L \ln \frac{E_t}{E_L}$$

which is derived by the relation:

$$E_t=E_L\epsilon^{-\frac{W_L}{\tau_L}}$$

Thus, the average output voltage is a logarithmic function of the input voltage E_t. In the antilogarithmic circuit, a transistor switch is used to modulate the width of the exponential pulses e_L that will appear on the load. The input information is contained in the pulse width W_L of a train of rectangular pulses (not shown in Fig. 13) which are used to modulate the on-off time of the transistor switch. The average output is, therefore, an antilogarithmic function of the input pulse width W_L as shown in the following equation:

$$E_0(\text{avg})\cong f\int_{W_L}^{\infty} E_L\epsilon^{-\frac{t}{\tau_L}}\,dt$$

$$=E_L\tau_L f\,\text{antiln}\left(-\frac{W_L}{\tau_L}\right)$$

Other transcendental functions can be obtained by using the same principle if the exponential pulses are replaced by pulses of different wave forms. A detailed description of our circuits is intended to be published in the future.

D. H. Schaefer and R. L. Van Allen: We are happy to see alternate methods of performing analogue computations. The experimental results obtained with these circuits will be of great interest.

Synthesis of Transfer Functions with Poles Restricted to the Negative Real Axis into Two Parallel R-C Ladders and an Ideal Transformer

M. G. MALTI
MEMBER AIEE

HUN HSUAN SUN
ASSOCIATE MEMBER AIEE

Synopsis: This paper presents a method of synthesizing a circuit to obtain a transfer function whose poles lie on the negative real axis and whose zeros lie anywhere in the complex frequency plane including the positive real axis by means of two parallel resistance-capacitance (R-C) ladders and an ideal transformer. The polynomial in the numerator of the given transfer function is first resolved into the difference of two component polynomials whose zeros are all negative real numbers. The method of resolution is based on conditions for collinearity of points that represent *n*th degree polynomials (with real coefficients) in *n*-dimensional space. The two component polynomial fractions for the given transfer function are then synthesized into two R-C ladders and the resulting network is the parallel combination of these two ladders through an ideal transformer and terminated by another R-C ladder. The method applies to more generalized cases and usually gives a network that contains less resistive and capacitive elements than any of the other methods available. If the resolution of the numerator polynomial is as a sum of polynomials instead of a difference, more than two components may be required. An expression for the upper bound to the least number of such components is also included in the paper.

IN RECENT YEARS there has been considerable interest in the question of how a transfer function may be realized as a linear passive network containing a minimum number of resistances and capacitances. Methods have been derived so that a transfer function whose poles and zeros are all real and negative can be synthesized into an R-C unbalanced ladder.[1-3] Further developments have been made to extend these methods to the transfer function containing complex zeros, while the number of resistive and capacitive elements has been kept low by various methods of synthesis.[4-6]

Consider a rational fraction of the complex frequency variable p, as:

$$\frac{P(p)}{Q(p)}=\frac{p^n+a_1p^{n-1}+a_2p^{n-2}+\ldots+a_n}{p^m+b_1p^{m-1}+b_2p^{m-2}+\ldots b_m}$$

$$=\frac{(p-p_1)(p-p_2)\ldots(p-p_n)}{(p-q_1)(p-q_2)\ldots(p-q_m)}\quad n\leq m \quad (1)$$

where $p_1p_2\ldots p_n$ may be real and complex numbers, $q_1q_2\ldots q_m$ are negative real and simple. The coefficients of $P(p)$ are real numbers while the coefficients of $Q(p)$ are restricted to be positive and real as governed by the nature of their roots.[7] It is required to realize the rational fraction within a constant multiplier K as the transfer function of the ratio of the input and output voltages of an unbalanced R-C network, Fig. 1, as:

$$H(p)=\frac{E_2}{E_1}=K\frac{P(p)}{Q(p)} \qquad (2)$$

A method, already described in the literature,[4,6] is to realize the given rational function into a network containing a number of R-C ladders in parallel. The procedure is first to resolve the numerator polynomial $P(p)$ into a sum of component polynomials that contain only real negative zeros while the denominator polynomial $Q(p)$ is left unchanged. Each component polynomial together with the denominator polynomial is then synthesized into an R-C ladder. The sum of these component functions representing the transfer function is then realized by the parallel combination of these component ladders with the proper gain adjustment. This method applies only to the transfer function whose zeros lie on the left-hand side of the complex frequency plane and the number of parallel R-C ladders used has been found to be

$$t\leq\frac{n+1}{2}\quad\text{if } n \text{ is odd}$$

$$\leq\frac{n}{2}+1\quad\text{if } n \text{ is even} \qquad (3)$$

n is the degree of the numerator.

If the numerator polynomial $P(p)$ contains both negative real and complex zeros, the number of parallel R-C ladders can be reduced by first factoring the negative real zeros, as:

Paper 56-126, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted April 13, 1955; made available for printing November 21, 1955.

M. G. MALTI is with Cornell University, Ithaca, N. Y., and HUN HSUAN SUN is with the Drexel Institute of Technology, Philadelphia, Pa.

This paper is a condensation of Dr. Sun's Ph.D. Thesis, part I, at Cornell University.

The authors wish to express their gratitude to Dr. H. Sohon, University of Pennsylvania, for his valuable suggestions and encouragement; Dr. F. C. Morneghem, Institute Tech. de Aero, Est São Paulo, Brazil, for suggesting the collinear conditions of points represented by polynomials; and S. S. Abhyanker, Harvard University, for suggesting the proof of the uppermost bound for the least number of component polynomials required in the resolution of a polynomial.

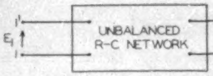


Fig. 1. Unbalanced R-C network

$$P(p^n)=P_1(p^{n-r})\times P_2(p^r) \quad (4)$$

where $P_2(p^r)$ is the polynomial formed by the r number of nonpositive real zeros and $P_1(p^{n-r})$ is the polynomial formed by the remaining zeros. The resolution of $P_1(p^{n-r})$ will therefore result in a smaller number of parallel ladders as:

$$\begin{aligned} t &\leqslant \frac{(n-r)+1}{2} && \text{if } (n-r) \text{ is odd} \\ &\leqslant \frac{(n-r)}{2}+1 && \text{if } (n-r) \text{ is even} \end{aligned} \quad (5)$$

Then by the application of Norton's network theorem,[6] the rational fraction formed by the zeros of $P_1(p^{n-r})$ together with a number of poles equal to one less than the number of zeros is synthesized into parallel R-C ladders, while the rational fraction formed by the zeros of $P_2(p^r)$ and the remaining poles is synthesized into an R-C termination ladder connected to the output of the parallel R-C ladders. If r is equal to zero, or the given transfer function contains no negative real zeros, then equation 5 reduces to equation 3 while the termination ladder becomes a unity resistance as originally derived by Guillemin.[4] The number of R-C elements used in each ladder has been further reduced by the use of a zero-cancellation device. This method has been extended to the transfer function containing zeros in the whole complex frequency plane excluding the positive real axis by the use of a surplus factor.[5] If the zeros of the given transfer function are close to the positive real axis, the degree of the equation which represents the surplus factor becomes extremely large. This, in turn, requires the addition of a large number of R-C elements.

The method to be presented in this paper, however, reduces the number of parallel R-C ladders used to only two without increasing the degree of the numerator polynomial of the given transfer function, and applies to the transfer function whose zeros lie in the whole complex frequency plane including the positive real axis. A method is developed to resolve any polynomial with real coefficients into the difference of only two component polynomials each containing negative real zeros. The synthesized network will, therefore, contain only two R-C ladders with their inputs in parallel, their output connected to an additional R-C ladder through an ideal transformer. The ideal transformer is necessary in order to avoid a circulating current.[8] A theorem is presented to show that the uppermost bound of the minimum number of component polynomials whose zeros are negative real and whose sum is equal to the given polynomial is the least integer greater than $n/2$, where n is the degree of the given polynomial.

A numerical example is given to show the method of synthesizing a transfer function with zeros unrestricted and poles on the negative real axis of the complex frequency plane into two parallel R-C ladders and an ideal transformer.

Resolution of a Polynomial with Real Coefficients into the Difference of Two Component Polynomials Each with Real Negative Zeros

It has been found that a polynomial with real coefficients in a single variable can be represented as a point in a geometrical space.[9] Therefore, we can say that a polynomial of

$$M(x)=x^n+a_1x^{n-1}+a_2x^{n-2}+\ldots+a_n \quad (6)$$

where $a_1a_2 \ldots a_n$ are real numbers, is a point defined by its coefficients as co-ordinates in an n-dimensional space represented by:

$$N(a_1a_2\ldots a_n) \quad (7)$$

The coefficients of the highest degree term is adjusted to be unity. A polynomial of the form

$$R(x)=x^n \quad (8)$$

is thus the point that coincides with the origin.

A polynomial with positive real coefficients can therefore be regarded as a point defined by its coefficients as co-ordinates in an n-dimensional space that lies in a domain whose co-ordinates are all positive. Such a domain shall be referred to as the positive domain. The n-dimensional space can be considered as divided into 2^n separate domains; for example, if $n=2$, the domains are quadrants, if $n=3$, the domains are octants, and so forth.

A discriminant of a polynomial is the equation that determines whether the given polynomial contains real zeros or not. Thus, from the foregoing discussion, a discriminant shall be defined as the boundary between the real-root domain and the complex-root domain. The discriminants of polynomials lower than fourth degree have been determined[7] and can be expressed in terms of the coefficients of the given polynomials. There are various methods available to determine whether an nth degree polynomial contains real zeros or not, but a unique expression in terms of the coefficients of the given polynomial that satisfies the foregoing condition is necessary in order to be regarded as a discriminant. Appendix I gives an expression in the form of determinants that serves as a criterion for negative real zeros of a given nth degree polynomial and can thus be regarded as the discriminant of the polynomial.

This basis will be used to distinguish the polynomials that contain only real zeros from those containing complex zeros. The space where the polynomial is represented is separated into two domains, namely, the real-root domain and the complex-root domain. The real-root domain is defined as that portion of the n-dimensional space which contains all the points $N(a_1a_2 \ldots a_n)$ that make the polynomial $M(x)=x^n+a_1x^{n-1}+\ldots+a_n$ have only real zeros. Similarly, the complex-root domain shall be defined as that portion of the n-dimensional space which contains the points $N(a_1a_2 \ldots a_n)$ that make the polynomial $M(x)=x^n+a_1x^{n-1}+\ldots+a_n$ have one or more pairs of imaginary or complex zeros. It is, therefore, evident that those points represented by $N(a_1a_2 \ldots a_n)$ which are not in the real-root domain must lie in the complex-root domain.

The problem in R-C network synthesis is to express a polynomial that lies in the complex-root domain as a linear combination of as few as possible component polynomials that lie in a domain common to the positive domain and real-root domain. The former condition is required by the fact that the coefficients of a polynomial with negative real zeros must be positive.[7] Now consider the following three polynomials:

$$P(x)=x^n+a_1x^{n-1}+a_2x^{n-2}+\ldots+a_n \quad (9)$$

$$B(x)=x^n+b_1x^{n-1}+b_2x^{n-2}+\ldots+b_n \quad (10)$$

$$C(x)=x^n+c_1x^{n-1}+c_2x^{n-2}+\ldots+c_n \quad (11)$$

where $P(x)$ is the given polynomial and $B(x)$ and $C(x)$ are its component polynomials. From these discussions, they can be represented by the following three points:

$$G(a_1a_2\ldots a_n) \quad (12)$$

$$H(b_1b_2\ldots b_n) \quad (13)$$

$$F(c_1c_2\ldots c_n) \quad (14)$$

If the polynomials of equations 9, 10, and 11 are linearly dependent, then their points represented by equations 12, 13, and 14 must be collinear.[9] As three points are the least number of points for the collinear condition in n-dimensional space, therefore, the minimum number of component polynomials has been chosen as two. By a well-known theorem of linear dependence,[9] it must be true that every 3-rowed determinant of the following matrix

$$\left\|\begin{matrix} 1 & a_1 & a_2 & \cdots & a_n \\ 1 & b_1 & b_2 & \cdots & b_n \\ 1 & c_1 & c_2 & \cdots & c_n \end{matrix}\right\| \tag{15}$$

must vanish, or the following determinant must be equal to zero, that is,

$$\begin{vmatrix} 1 & a_i & a_j \\ 1 & b_i & b_j \\ 1 & c_i & c_j \end{vmatrix} = 0 \tag{16}$$

where i and j are both positive integers such that $1 \leqslant i < n$, $1 \leqslant j \leqslant n$ and $i \neq j$.

Expanding equation 16 on the last column and equating to zero gives:

$$\begin{vmatrix} 1 & a_i \\ 1 & c_i \end{vmatrix} b_j + \begin{vmatrix} 1 & b_i \\ 1 & a_i \end{vmatrix} c_j = \begin{vmatrix} 1 & b_i \\ 1 & c_i \end{vmatrix} a_j$$

or

$$k_1 b_j + k_2 c_j = a_j \tag{17}$$

where

$$k_1 = \frac{c_i - a_i}{c_i - b_i} \quad \text{if } c_i \neq b_i \tag{18}$$

$$k_2 = \frac{a_i - b_i}{c_i - b_i} \quad \text{if } c_i \neq b_i \tag{19}$$

and from equations 18 and 19:

$$k_1 + k_2 = 1 \tag{20}$$

or

$$k_1 b_j x^{n-j} + (1 - k_1) c_j x^{n-j} = a_j x^{n-j} \qquad j = 1 \ldots n \tag{21}$$

Thus, the following relationship between the given polynomial $P(x)$ and its two component polynomials $B(x)$ and $C(x)$ are arrived at, as:

$$P(x) = k_1 B(x) + (1 - k_1) C(x) \tag{22}$$

where k_1 is given by equation 18.

As an illustration, consider the following cases:

1. Second-degree polynomials. Let

$$P(x) = x^2 + a_1 x + a_2 \tag{23}$$

be the given polynomial with positive real coefficients and complex zeros. For this simple case, the discriminant is

$$a_1^2 - 4a_2 = 0 \tag{24}$$

and can be represented as a parabola in the 2-dimensional space shown in Fig. 2. Point $G(a_1a_2)$ in Fig. 2 represents the given polynomial (equation 23), and must lie on the left-hand side of the discriminent and in the first quadrant. Draw any straight line from $G(a_1a_2)$ and select points $H(b_1b_2)$ and $F(c_1c_2)$. The points H and F thus will be collinear with G and must be selected to lie on the right-hand side of the discriminant and in the first quadrant.

Draw the vector **OG**, **OH**, and **OF** from the origin to the three points G, H, and F, respectively. Since G, H, and F are collinear, then vectors **FG** and **FH** must have the same direction and their ratio λ is a real number which will be shown later to be k_1.

$$\mathbf{FG} = \lambda \mathbf{FH} \tag{25}$$

As H lies between G and F as shown in Fig. 2, therefore λ is positive real and greater than one. Also it is seen that

$$\mathbf{OF} - \mathbf{OG} = \lambda[\mathbf{OF} - \mathbf{OH}] \tag{26}$$

Solving for **OG**, a relationship is obtained similar to equation 22:

$$\mathbf{OG} = \lambda \mathbf{OH} + (1 - \lambda)\mathbf{OF} \tag{27}$$

Let the termini of the vectors **OG**, **OH**, and **OF** represent respectively the points whose co-ordinates are the coefficients of the polynomials $P(x)$, $B(x)$, and $C(x)$ in the 2-dimensional space, it then can be stated that:

$$P(x) = \lambda B(x) + (1 - \lambda) C(x) \tag{28}$$

where

$$B(x) = x^2 + b_1 x + b_2 \tag{29}$$

and

$$C(x) = x^2 + c_1 x + c_2 \tag{30}$$

It is obvious that $P(x)$, $B(x)$, and $C(x)$ as shown are linearly dependent and that λ is identical with k_1. Since λ is greater than unity, the original polynomial has been resolved into the difference of two polynomials.

However, if the point G is chosen so that it lies between H and F in the collinear condition shown in Fig. 3, then, as before,

$$\mathbf{FG} = \lambda \mathbf{FH} \tag{31}$$

where λ is positive, real, and less than one as can be seen from Fig. 3. Thus the original polynomials become the sum of two polynomials

$$\mathbf{OG} = \lambda \mathbf{OH} + (1 - \lambda)\mathbf{OF} \tag{32}$$

where λ is likewise equal to k_1.

If G is between H and F and is very close to the a_2-axis, then the distance between H and F becomes very large. In the limiting case H would be chosen at the origin, G on the a_2-axis, and F will be the other end of the straight line OGF that terminates in the real-root domain at infinity. Since the results are quite simple, it is no problem to resolve a second-degree polynomial into a sum or difference of two polynomials having real negative zeros.

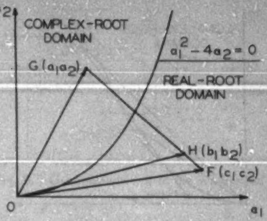


Fig. 2. Resolution of a second-degree polynomial into the difference of two polynomials

2. Third-degree polynomial. Let

$$P(x) = x^3 + a_1 x^2 + a_2 x + a_3 \tag{33}$$

be the given polynomial with positive real coefficients and complex zeros. The discriminant of equation 33 can be represented as a surface in 3-dimensional space. Assuming that G lies in the complex-root domain, a straight line can always be drawn from G that terminates in the positive octant and in the real-root domain. Point H is then selected on the line in the real-root domain. Point F which is also in the real-root domain is obtained by extending the straight line. The procedure followed is then the same as previously outlined. The plot for the boundary is not included for the third degree as it is not a simple surface.

3. Nth degree polynomials. From the foregoing discussion, it can be seen that a polynomial represented by a point in the complex-root domain can always be expressed as the linear combination of two polynomials represented by their point in the real-root domain. However, due to the complexity of the boundary for higher degree polynomials, it may be difficult to have G located between H and F, whereas it is much easier to select F near H and determine the distance between F and H. Thus, the resolution of $P(x)$ is the difference of two polynomials having negative real zeros with k_1 greater than one.

The foregoing method can easily be extended to polynomials having positive real zeros or zeros with positive real parts. This will be sufficient to introduce negative coefficients in the given polynomial. Consider again the following second-degree polynomial:

$$P(x) = x^2 + a_1 x + a_2 \tag{34}$$

where a_1 is a negative real number and a_2 is a positive real number. The point

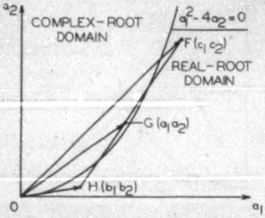


Fig. 3. Resolution of a second-degree polynomial into the sum of two polynomials

$G(a_1a_2)$ as representing equation 34 is shown in Fig. 4. The collinear condition still exists, hence two polynomials are suffiicent. This case can thus be extended to the n-dimensional case, if it is remembered that a straight line can always pass from any point in the complex-root domain to a point in the real-root domain of positive coefficients.

Detailed Procedures for Resolving a Polynomial with Real Coefficients into the Difference of Two Polynomials with Negative Real Zeros

If $B(x)$ and $C(x)$ of equations 10 and 11 are represented by points H and F of equations 13 and 14 in n-dimensional space and lie near to each other, then it can be said that:

$$c_i=\mu_i b_i \qquad i=1, 2, \ldots n \tag{35}$$

Since polynomials $B(x)$ and $C(x)$ are to have negative real zeros, their coefficients are necessary positive numbers. Hence

$$\mu_i>0 \tag{36}$$

Substituting equation 35 in 18 gives

$$k_1=\frac{\mu_i-\dfrac{a_i}{b_i}}{\mu_i-1} \tag{37}$$

$$=\frac{\beta_i\mu_i-1}{(\mu_i-1)\beta_i} \tag{38}$$

where

$$\beta_i=b_i/a_i \qquad i=1, 2, \ldots n \tag{39}$$

and

$$\mu_i\neq 1 \tag{40}$$

As has been shown, $\mu_i\neq 1$, except when $a_i/b_i=1$. In this case $k_1=1$ or the polynomials $P(x)$ and $B(x)$ are identical as will be discussed later.

Therefore there must be constructed a polynomial $B(x)$ which has the same degree as the given polynomial, and whose point $H(b_1b_2 \ldots b_n)$ must lie in the domain common to the positive and real-root domains. After $B(x)$ is obtained, then a value of $\mu_s>0$ must be assumed for a particular set of a_s, b_s, and c_s. k_1 is then determined for this assumed value of μ_s. To simplify this process, Fig. 5 was plotted for β_s against k_1 with μ_s as parameter for the cases of positive value of $\beta_s (s=1 \ldots n)$. Fig. 6 was plotted for the case of negative value of β_s, $(s=1 \ldots n)$ or the given polynomial contains negative coefficients.

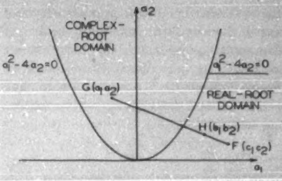


Fig. 4. Resolution of a second-degree polynomial containing negative coefficients

From the discussion, a procedure can be arrived at which can be easily followed in the resolution process. To simplify the procedure, the following is chosen:

$$b_n=c_n=a_n \tag{41}$$

Geometrically, this implies that G, H, and F are all on the straight line perpendicular to the a_n-axis of the n-dimensional space.

1. Construct a polynomial.

$$B(x)=(x+r_1)(x+r_2)\ldots(x+r_n) \tag{42}$$

$$=x^n+b_1x^{n-1}+\ldots+b_{n-1}x+a_n \tag{43}$$

where $r_1r_2 \ldots r_n$ are positive real numbers such that the resulting polynomial, equation 43, will have its coefficients satisfy the following equation:

$$b_i>a_i \qquad i=1\ldots n-1 \tag{44}$$

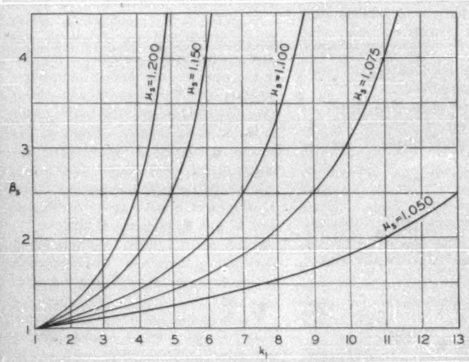


Fig. 5. β_s **versus** k_1 **(positive value of** β_s**)**

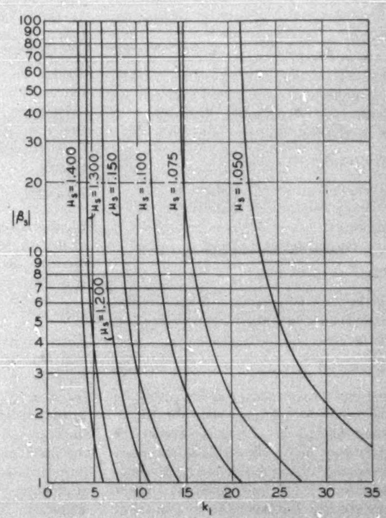


Fig. 6 (right). β_s **versus** k_1 **(negative values of** β_s**)**

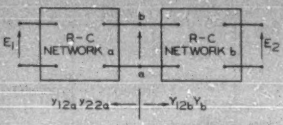


Fig. 7. Two R-C networks in cascade

2. Find β_i from equation 39.

3. Choose one β_i say β_s and assume a value of μ_s. To choose μ_s, Fig. 5 or 6 can be used as a first guess. By the chosen value of β_s, a rough estimate of the relative magnitude of μ_s and k_1. can be obtained. Thus

$$k_1=\frac{\beta_s\mu_s-1}{(\mu_s-1)\beta_s} \tag{45}$$

4. After k_1 is obtained, then μ_j can be found through the relation:

$$\mu_j=\frac{\beta_jk_1-1}{\beta_j(k_1-1)} \qquad \begin{matrix} j=1\dots n-1 \\ j\neq s \end{matrix} \tag{46}$$

5. The coefficients of $C(x)$ can be obtained by using the relation

$$c_i=\mu_ib_i \qquad i=1\dots n-1 \tag{47}$$

6. Test $C(x)$ by the real-root criterion as outlined in Appendix I. If it is not satisfied, reduce μ_s and proceed again.

The Uppermost Bound of the Minimum Number of Polynomial Whose Sum Is Equal to a Given Polynomial

The foregoing method gives a powerful tool for resolving a polynomial into polynomials having negative real zeros. In general, it has been shown[10] that the minimum number t of the components of a polynomial with positive real coefficients of nth degree is between the lowest bound of two and the uppermost bound which is the least integer greater than $n/2$ or

$$\begin{aligned} 2\leqslant t &\leqslant\frac{n+1}{2} && \text{if } n \text{ is odd} \\ &\leqslant\frac{n}{2}+1 && \text{if } n \text{ is even} \end{aligned} \tag{48}$$

To show that the uppermost bound is that given in equation 48 consider the polynomial having positive real coefficients as:

$$P(x)=x^n+\alpha_1x^{n-1}+\delta_2x^{n-2}+\alpha_3x^{n-3}+\delta_4x^{n-4}+\dots+\delta_n \tag{49}$$

where n is even and α's are very small compared with δ's or

$$\delta_{i+1}>>\alpha_i \qquad i=1, 2, \dots n-1 \tag{50}$$

In case n is odd, the real zero can always be factored out and a polynomial of even degree obtained.

It is always possible to write $P(x)$ as

$$P(x)=Q_1(x)+\sum_{p=1}^{n/2} C_pR_p(x) \tag{51}$$

where $Q_1(x)$ is an nth degree polynomial with all zeros real and negative and $C_p>0$. Since α_1 is very small and the coefficients of all $R_p(x)$, $p=1 \dots n/2$, are all positive real, therefore the coefficients of $Q_1(x)$ must be all very small. This implies that the zeros of $Q_1(x)$ are very small.[7] Let $Q_1(x)$ be expressed as:

$$Q_1(x)=x^n+d_1x^{n-1}+d_2x^{n-2}+\dots+d_n \tag{52}$$

and let

$$d_1=\alpha_1 \tag{53}$$

and

$$d_j\leqq\alpha_j \qquad j=2, 4, \dots n \tag{54}$$

This can only be obtained by choosing

$$\epsilon_i>0 \qquad i=1, 2, \dots n \tag{55}$$

small enough such that

$$\epsilon_1+\epsilon_2+\dots+\epsilon_{n-1}=\epsilon_n<\alpha_1 \tag{56}$$

then

$$Q_1(x)=(x+\epsilon_1)(x+\epsilon_2)\dots[x+(\alpha_1-\epsilon_n)] \tag{57}$$

has negative real zeros.

Since $\alpha_1=d_1$, then $P(x)-Q_1(x)=\sum_{p=1}^{n/2}C_p R_p(x)$ contains no terms of degree higher than $n-2$. The degree has thus been reduced by two in the resolution. The degree cannot be reduced more than two at a time, since the next coefficients δ_2 of the given polynomial is much larger than α_1, hence $Q_1(x)$ cannot be obtained to carry off the first three terms of $P(x)$ without sacrificing the condition of its negative real zeros.

By repeating this process, at each step a new polynomial is obtained of two degrees lower than before. Thus the minimum number is $n/2+1$.

However, a polynomial could always be resolved simply by grouping the successive terms two at a time, as:

$$P(x)=(x+a_1x^{n-1})+(a_2x^{n-2}+a_3x^{n-3})+\dots+a_n \tag{58}$$

The number of resolution is $(n+1)/2$ if n is odd, and $n/2+1$ if n is even. We do not want to resort to this kind of resolution, as better results may be obtained by other means, unless the given polynomial is of the type mentioned in equation 49. Appendix II gives the theorem that defines the condition under which the given polynomial can only have this minimum number of components.

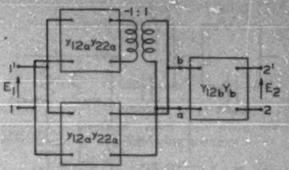


Fig. 8. Generalized form of the synthesized network

Example

The synthesis presented is based on the foregoing method of resolving the numerator polynomial into the difference of two component polynomials containing negative real zeros. The zero-cancellation device and the application of Norton's network theorem as derived by Weinberg,[5] together with the zero-shifting method as derived by Guillemin[8] are used in the example.

From equations 2 and 4 the transfer function is

$$H(p)=K\frac{P_1(p)P_2(p)}{Q(p)} \tag{59}$$

Introduce

$$W(p)=W_1(p)xW_2(p) \tag{60}$$

such that

$$H(p)=K\frac{\left(\frac{P_1(p)}{W_1(p)}\right)\left(\frac{P_2(p)}{W_2(p)}\right)}{\frac{Q(p)}{W_1(p)W_2(p)}} \tag{61}$$

The denominator of equation 61 can be partial-fractioned as

$$=K\frac{\left(\frac{P_1(p)}{W_1(p)}\right)\left(\frac{P_2(p)}{W_2(p)}\right)}{\frac{Q_1(p)}{W_1(p)}+\frac{Q_2(p)}{W_2(p)}} \tag{62}$$

$$=K\frac{(y_{12a}(p))(Y_{12b}(p))}{y_{22a}(p)+Y_b(p)} \tag{63}$$

where $y_{12a}(p)$ and $y_{22a}(p)$ are the short-circuit transfer and driving point admittances of network a and $Y_{12b}(p)$ and $Y_b(p)$ are the short-circuit transfer and driving point admittances of network b as shown in Fig. 7. Network b is the termination ladder and network a contains the parallel combination of the two R-C ladders and an ideal transformer.

Now resolve $P_1(p)$ by the method outlined before:

$$P_1(p)=k_1P_1^{(1)}(p)+k_2P_1^{(2)}(p) \tag{64}$$

where $k_2=1-k_1$. Then equation 63 becomes

$$H(p)=\frac{[k_1y_{12a}^{(1)}+k_2y_{12a}^{(2)}]\,Y_{12b}}{y_{22a}+Y_b} \tag{65}$$

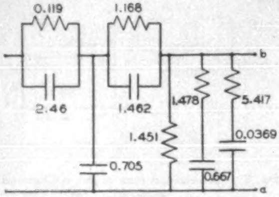


Fig. 9. Example: first parallel ladder

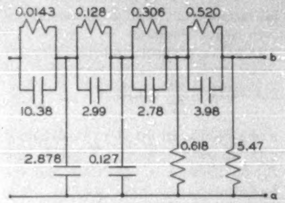


Fig. 10. Example: second parallel ladder

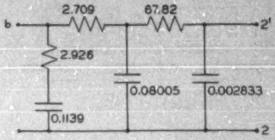


Fig. 11. Example: termination ladder

The factoring of $W(p)$ into $W_1(p)$ and $W_2(p)$ must be made such that $W_1(p)$ is one degree lower than $P_1^{(1)}(p)$ or $P_1^{(2)}(p)$ in order to be physically realizable.[4,8,11]

To reduce the number of elements, the process of resolving $P_1(p)$ and the factoring of $W(p)$ should be co-ordinated so that zero-cancellation may occur. This may be done by factoring the $W(p)$ in such a way that the component $W_1(p)$ contains some of the zeros of $P_1^{(1)}(p)$ or $P_1^{(2)}(p)$ and the component $W_2(p)$ contains some of the zeros of $P_2(p)$.

From equation 63, $y_{12a}^{(1)}y_{22a}$ and $y_{12a}^{(2)}y_{22a}$ are each synthesized from right left into two R-C ladders by the zero-shifting method. They are then paralleled to through an ideal transformer of $-1:1$ ratio as shown in Fig. 8 after the proper gain adjustment. To adjust the gain, the following can be done:

There is given

$$y_{12a}=k_1y_{12a}^{(1)}+k_2y_{12a}^{(2)} \tag{66}$$

After the synthesis,

$$(A_1y_{12a}^{(1)}+A_2y_{12a}^{(2)}) \tag{67}$$

A_2 here is assumed to be negative as this includes the ideal transformer of $-1:1$ ratio and $k_2=1-k_1$. In order to bring the y_{22a} of the two ladders to the same value, the gain of y_{12a} must be adjusted. Hence, write the given y_{12a} as

$$y_{12a}=\frac{k_1}{A_1}[A_1y_{12a}^{(1)}]+\frac{k_2}{A_2}[A_2y_{12a}^{(2)}] \tag{68}$$

Therefore the over-all y_{12a} of the two ladders and the ideal transformer is changed by a factor of

$$G=\frac{k_1}{A_1}+\frac{k_2}{A_2} \tag{69}$$

Then each component $y_{12a}^{(1)}$ of equation 66 will be multiplied by the corresponding factor of

$$\frac{k_i}{A_iG} \qquad i=1,\ 2 \tag{70}$$

Now consider the following example.

Given the transfer function:

$$H(p)=K\frac{(p+3)(p^4-2p^3-p^2+2p+10)}{(p+0.5)(p+1.5)(p+2.5)(p+3.5)(p+4.5)(p+5.5)(p+6.5)}$$

it is required to synthesize $H(p)$ into two parallel R-C ladders and ideal transformer with termination ladder within a constant multiplier. The procedure can be done by the following steps.

1. The numerator can be represented:

$$P_1(p)=p^4-2p^3-p^2+2p+10$$
$$P_2(p)=p+3$$

where $P_1(p)$ and $P_2(p)$ are defined in equation 4.

2. By use of the resolution procedures as outlined before, $P_1(p)$ can be resolved as follows:

(*a*) Construct $P_1^{(1)}(p)=(p+0.5858)(p+1)(p+3.4142)(p+5)$ $=p^4+10p^3+31p^2+32p+10$

(*b*) $\beta_1=-5,\ \beta_2=-31,\ \beta_3=16$

(*c*) Choose $\beta_3=\beta_1=-5$. From Fig. 6, choose $\mu_3=1.1,\ k_1=13,\ \mu_3=1.075,\ k_1=16$;

Use

$$\mu_3=1.1,\ k_1=13$$

Compute

(*d*) $\mu_2=1.086,\ \mu_3=1.078$

(*e*) $c_1=11,\ c_2=33.67,\ c_3=34.50$

(*f*) $P_1^{(2)}(p)=p^4+11p^3+33.67p^2+34.50p+10$;

$P_1^{(2)}(p)$ in item (*f*) satisfies the reality criterion of equation 71. The following equation is thus obtained:

$$\begin{aligned}P_1(p)&=13(p^4+10p^3+31p^2+32p+10)-12(p^4+11p^3+33.667p^2+34.50p+10)\\&=13(p+0.5858)(p+1)(p+3.4142)(p+5)-12(p+0.4830)(p+1.175)(p+2.619)(p+6.723)\end{aligned}$$

3. Choose $W(p)$

$$W(p)=(p+1)(p+2)(p+3)(p+4)(p+5)(p+6)$$

As can be seen, $W(p)$ has one zero-cancellation with $P_2(p)$ and two zero-cancellations with $P_1^{(1)}(p)$. Hence, factoring $W(p)$ into $W_1(p)$ and $W_2(p)$ results in the following:

$$W_1(p)=(p+1)(p+2)(p+5)$$
$$W_2(p)=(p+3)(p+4)(p+6)$$

4. All the component functions can thus be written according to equations, 64, 63, and 61 as:

$$y_{12a}^{(1)}=\frac{P_1^{(1)}(p)}{W_1(p)}=\frac{(p+0.5858)(p+3.4142)}{(p+2)}$$

$$y_{12a}^{(2)}=\frac{P_1^{(2)}(p)}{W_1(p)}=$$

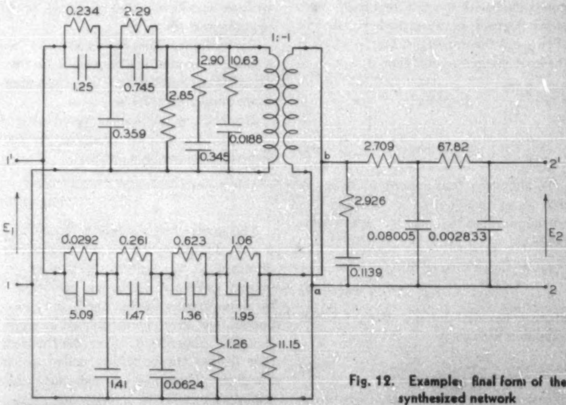


Fig. 12. Example: final form of the synthesized network

$$\frac{(p+0.4830)(p+1.175)(p+2.619)(p+6.723)}{(p+1)(p+2)(p+5)}$$

$$y_{21a}=\frac{Q_1(p)}{W_1(p)}=(p+1.4663)+\frac{0.6768p}{p+1}+\frac{0.4613p}{p+2}+\frac{0.1846p}{p+5}$$

$$y_{22a}=\frac{p^4+10.789p^3+36.790p^2+44.341p+14.663}{p^3+8p^2+17p+10}$$

$$Y_{12b}=\frac{P_2(p)}{W_2(p)}=\frac{1}{(p+4)(p+6)}$$

$$Y_b=\frac{Q_2(p)}{W_2(p)}=\frac{0.3418p}{p+3}+\frac{0.2564p}{p+4}+\frac{0.1128p}{p+6}$$

5. Synthesize network a from right to left. First synthesize $y_{12a}^{(1)}$ and y_{22a} by the zero-shifting method into an R-C ladder. The poles at $p=-1$ and $p=-5$ must be removed first from y_{22a} as parallel branches and the resulting R-C ladder is shown in Fig. 9. Now synthesize $y_{12a}^{(2)}$ and y_{22a} and the resulting R-C ladder is shown in Fig. 10.

6. Synthesize the Y_{12b} and Y_b as the termination ladder of network b from left to right. The shunt branch $\frac{0.3418p}{p+3}$ is first removed and the remaining terms can be synthesized by the method of continued fractions.[4] The resulting ladder is shown in Fig. 11.

7. Adjust the gain of the two ladders of network a and then parallel them through the ideal transformer. As $k_1=13$ and $k_2=-12$, and, from the synthesized ladders, $A_1=0.7764$ and $A_2=-0.7446$, thus:

$$G=\frac{13}{0.7764}+\frac{12}{0.7746}=32.86$$

Hence, for the first ladder in network a, each R should be divided by $A_1G_1/k_1=0.510$ and each C should be multiplied by 0.510. For the second ladder in network a each R should be divided by $k_2/A_2G_2=0.491$, and each C should be multiplied by 0.491.

The final network is then shown in Fig. 12. It has a total of only 28 elements.

Appendix I Criterion for the Negative Reality of Zeros of a Polynomial[10]

Theorem I

If $P(x)=a_0x^n+a_1x^{n-1}+a_2x^{n-2}+\ldots+a_n$ is a polynomial with all coefficients real and positive, then a necessary and sufficient condition that the zeros of the polynomial be real, distinct, and negative is that

$$S_p>0 \qquad p=1, 2, \ldots n \tag{71}$$

where S_1, S_2, $\ldots S_p$, $\ldots S_n$ are the first, third, $\ldots$ and $(2n-1)$th principal minors (blocked off by lines) in the array

$$\left\|\begin{array}{cccccc} a_1 & 2a_2 & 3a_3 & 4a_4 & 5a_5 & 6a_6\ldots \\ a_0 & a_1 & a_2 & a_3 & a_4 & a_5\ \ldots \\ 0 & a_1 & 2a_2 & 3a_3 & 4a_4 & 5a_5\ldots \\ 0 & a_0 & a_1 & a_2 & a_3 & a_4\ \ldots \\ 0 & 0 & a_1 & 2a_2 & 3a_3 & 4a_4\ldots \\ 0 & 0 & a_0 & a_1 & a_2 & a_3\ \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \multicolumn{6}{l}{a_j=0 \text{ if } j>n} \end{array}\right\| \tag{72}$$

The zeros of $P(x)$ must be distinct, hence if there are repeated zeros present, they should be factored out first by the Euclidean algorithm of the H.C.F. process between $P(x)$ and $P'(x)$.[7]

Appendix II Condition for Uppermost Bound of Minimum Number of Components in the Resolution of a Polynomial[10]

Theorem II

If $m(n)$ be the least integer such that any nth degree polynomial with positive coefficients can be expressed as a linear combination with positive coefficients of at most $m(n)$ polynomials, each of which has all its zeros negative, then $m(n)$ is the least number greater than $n/2$.

The relative magnitude of α's and δ's of equation 49 that should be attained when the minimum number of components in $n/2+1$ will be

$$P(x)=x^n+a_1x^{n-1}+\left\{\binom{n}{2}\binom{na_1}{2}^2+k_2\right\}x^{n-2}+a_3x^{n-3}+\left[\binom{n}{4}\left\{\binom{na_1}{2}^4+a_2\left[\frac{(n-2)\binom{n}{2}a_2}{2K_2}\right]^2\right\}+k_4\right]x^{n-4}+a_5x^{n-5}+\ldots \tag{73}$$

where a_i is the coefficient of x^{n-i} in $P(x)$ and is determined after $a_1\ldots a_{i-1}$ are determined in the following manner

$$a_i=\begin{cases}\text{arbitrary} & \text{if } i \text{ is odd}\\ \binom{n}{i}\sum_{j=0}^{1/2}a_{2j}\times\left[\frac{(n-2j)\binom{n}{2j}a_{2j}+1}{2k_{2j}}\right]^{i-2j}+k_i & \text{if } i \text{ is even}\end{cases}$$

$$a_0=k_0=1 \tag{74}$$

and k_i is an arbitrary positive number which occurs in the formula for a_i.

References

1. R-C Coupling Network, E. A. Guillemin. Radiation Laboratory Report, Massachusetts Institute of Technology, Cambridge, Mass., no. 43, Oct. 11, 1944.

2. The Synthesis of R-C Network to Have Prescribed Transfer Functions, H. J. Orchard. *Proceedings*, Institute of Radio Engineers, New York, N. Y., Apr. 1951, pp. 428–32.

3. A Note on the Ladder Development of R-C Networks, E. A. Guillemin. *Ibid.*, Apr. 1, 1952, pp. 482–85.

4. Synthesis of R-C Networks, E. A. Guillemin. *Journal of Mathematics and Physics*, Massachusetts Institute of Technology, Cambridge, Mass., vol. 28, no. 1, Apr. 1949, pp. 22–44.

5. Synthesis of Transfer Function with Poles Restricted to the Negative Real Axis, L. Weinberg. *Journal of Applied Physics*, New York, N. Y., vol. 24, no. 2, Feb. 1953, pp. 207–16.

6. Synthesis of Paralleled 3-Terminal R-C Networks to Provide Complex Zeros in the Transfer Function, P. F. Ordung, G. S. Axelby, H. L. Krauss, W. P. Yetter. *AIEE Transactions*, vol. 70, pt. II, Oct. 1951, pp. 1861–67.

7. Theory of Equations (book), H. W. Turnbull. Oliver and Boyd, London, England, 1952, pp. 34–40, 81–107, 117–39.

8. Communication Networks (book), E. A. Guillemin. John Wiley and Sons, Inc., New York, N. Y., vol. 2, 1935, pp. 132–221.

9. Introduction to Higher Algebra (book), M. Bocher. The Macmillan Company, New York, N. Y., 1907, pp. 8–52.

10. R-C Network Synthesis, Hun Hsuan Sun. *Ph.D. Thesis*, Cornell University, Ithaca, N. Y., 1955.

11. A Summary of Modern Methods of Network Synthesis, E. A. Guillemin. "Advance, in Electronics," Academic Press, New Yorks N. Y., vol. 3, 1951, pp. 261–303.

◆

Discussion

I. Horowitz (Polytechnic Institute of Brooklyn, Brooklyn, N. Y.): Aside from the question of the desirability of ideal transformers in R-C networks, the basic problem considered by the authors is: How can a polynomial $P(p)$ be written as the difference of two polynomials, each with negative real zeros only. The following procedure suffices:

1. Divide $P(p)$ by $D(p)=(p+a_1)(p+a_2)\ldots(p+a_n)$, where n is at least the degree of P and the a_j's are real, positive, and distinct.

2. Make a partial fraction expansion of $P(p)/D(p)$. All residues are real. Collect all positive residue terms as F_1, and negative residue terms as $-F_2$.

$$\frac{P(p)}{D(p)}=F_1(p)-F_2(p)$$

3. Multiply both sides of this equation by $D(p)$, giving $P(p)=D(p)F_1(p)-D(p)F_2(p)$.

DF_1 and DF_2 are each polynomials with negative real zeros only. F_1 and F_2 are each R-C driving point impedances, hence have only negative real zeros. $D(p)$ includes factors cancelling all poles of F_1 and F_2. This problem arose in the design of active R-C networks.

Automatic Network Analysis with a Digital Computation System

S. Y. WONG
ASSOCIATE MEMBER AIEE

M. KOCHEN
NONMEMBER AIEE

THE ANALYSIS of large electric and fluid networks is of great importance in engineering. Hand computation is generally impractical when the network is large. The network analyzer is currently the only practical means for solving such problems.

Very few network analyzers are equipped with programmed input-output systems, so that much time is spent in the setting up of problems on them. Attempts to overcome some of the limitations of network analyzers by the use of digital computing machines for network calculations have been made.[1] The methods so far developed are intended mostly for small accounting machines of the punched-card variety. However, much manipulation of the problem is involved before it reaches the computing machine. Hence, digital methods have gained little favor among engineers in the field.

Mathematical methods are available to obtain solutions directly from the circuit diagram by inspection of the networks according to definite rules.[2,3] The necessity to scan the input repeatedly makes these methods ill-suited to small accounting machines. With the advent of large-scale high-speed electronic digital computers, the digital solution of network problems becomes practicable.

The Digital Computing System

The most important aspect of a digital computing system is that once a program for a particular network has been constructed, it can be stored in coded form suitable for machine input and for later use in analysis of modifications of the network. A master library serves to ensure that every program developed by any one digital computing system is available to every other interested system. If several such systems operate over a long period, it will be highly probable that a program will be available to handle most commonly encountered network engineering problems.

To determine whether a digital computing system can be set up with the maximum convenience to the user, a sample system was tested on the Institute for Advanced Study computer. One basic machine program for d-c networks was made, and several variations of this were used to solve different types of problems. The mathematical basis of this program and its range of applicability will be discussed, although any number of techniques for solving systems of linear equations, some preferable to the one employed here, could be substituted. The significant aspect of this paper is the procedure of the automatic formulation of the problem from a circuit diagram, rather than the details of solving the mesh and loop equations.

It is supposed that the problem originator provides the following information when he submits his problem to a computing system:

1. A circuit diagram with the resistances or admittances and the electromotive forces (emf) in every branch.
2. The unknown quantities to be computed; i.e. the currents in one or more branches, the potentials at certain nodes.

The following steps characterize the procedure within the computing system which was tested:

1. If the branches and/or junctions are not already numbered by the problem originator when he submits a circuit diagram, this task is done by the first member of the computing system who handles the problem. An algebraic description (an incidence matrix) of the circuit diagram is then prepared.
2. Transformation codes then convert these data into a form which is suitable for machine computation with one of the available coded programs. This can generally be accomplished in a matter of seconds.
3. The complete information about the circuit being thus stored in the machine, the general-purpose program to solve the network is also read into the machine, and the desired quantities are calculated and transformed back into a form which permits their ready interpretation.
4. These answers are then automatically printed in tabular form and/or entered back upon the circuit diagram.
5. The problem may be rerun to duplicate the answers, or these may be checked by comparison with certain known quantities.
6. The coded input information and the possible program variations are filed and stored (on cards, tape, etc.) in the system's library for possible repeat runs. It takes a matter of minutes to set up the machine for a run.

Network Topology

An electric network is a linear graph where the conductors are edges or branches and the junctions are vertices or nodes. Let the N nodes be numbered 0, $\ldots$, i, $\ldots$, N-1 in an arbitrary manner. In the sample problem which was solved, the following parameters were associated with the branch connecting the ith to the jth node.

Let Z_{ij} be the resistance, $a_{ij}=\dfrac{1}{Z_{ij}}$ be the admittance of that branch, and e be the absolute value of the emf of a battery in such a branch, where i is taken to be the number of the node connected to the negative terminal of the battery.

$$e_{ij}=\begin{array}{l} +e \text{ if } i=\text{Min}\,(i,j) \\ -e \text{ if } i=\text{Max}\,(i,j) \end{array} \quad \text{by convention}$$

It is evident that $e_{ij}=-e_{ji}$ and that $a_{ij}=a_{ji}$. By convention $e_{ii}\equiv 0$ and $a_{ii}\equiv 1$. v_i is the potential at the ith node relative to ground, which is taken as 0.

In order to represent the network topology with minimum data for the purpose of conserving memory space (eliminating the redundancy of writing the usual incidence matrix in which every branch appears twice) the following scheme of numbering the nodes was proposed: Suppose that the first i nodes have been numbered 0, ..., i-1 according to this scheme, and consider the ith node. Call the set of nodes which are connected to the ith node by a single branch, G_i. Include in G_i only those node numbers which are greater than i. It is clear that every combination $(i,i')\ i,i'=0,\ldots,N-1,\ i\neq i'$ will occur once and only once. Thus, the sequence of integers, grouped according to $G_0, G_1, \ldots, G_{N'}$, $N'\leq N-1$, specifies the network topology. Letting n be the

---

Paper 56-156, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted August 9, 1955; made available for printing December 15, 1955.

S. Y. WONG is with Philco Corporation, Philadelphia, Pa., and M. KOCHEN resides in Cambridge, Mass.

The research reported in this paper was supported jointly by the U. S. Army, Navy, and Air Force, and the Atomic Energy Commission under contract no. *DA-36-034-ORD-1646*, Project *TB3-0538* and contract no. *N-7-ONR-388(01)* with the Institute for Advanced Study.

The authors wish to thank W. S. Ku and C. Hoffman, Public Service Electric and Gas Company, Newark, N. J., for their co-operation, the Institute for Advanced Study for the use of the electronic computer, and H. H. Goldstine for his encouragement and advice.

number of branches, there are n integers, $\zeta_1, \ldots, \zeta_t, \ldots, \zeta_n$, representing destination nodes, in the foregoing sequence. The groups $G_0, G_1, \ldots, G_{N-1}$ can be stored in a sequential fashion in the memory. Since the integer t of G_t does not appear explicitly in the group G_t it is necessary to put this integer t ahead of the integers in G_t. It is convenient to associate a binary variable, η_t, called the nodal index, to every integer in this sequence. $\eta_t=1$ if the integer corresponds to t of G_t otherwise. The reason for requiring the number t to appear explicitly is that sometimes G_t may contain no integers, as in the case of a node boxed in by nodes of lower numbers.

It is then possible to record all the information necessary to specify the network topology in $(n+N)(\log_2(N)+1)$ binary digits, where (N) is the smallest power of 2 which is not less than N. The parameters a_{ij}, and e_{ij} are written next to every branch, and if these are known to an accuracy of A and $E+1$ bits, respectively, then the entire network can be described by a table which requires $(n+N)(A+E+2+\log_2(N))$ bits to store. To determine the incident node (with lower node number) for the tth branch, it is merely necessary to examine the number in G_t whose $\eta_t=1$. The destination node (with higher node number) is given by ζ_t. If $v_0, \ldots, v_{N-1}$ are stored in machine addresses numbered 0 to $N-1$, it is very convenient to let $i, \zeta_t=0, \ldots, N-1$ represent these addresses.

If a node has only two branches connected to it, the pair of branches represents two conductors in series, and such a node can be eliminated. If m $(m \geq 2)$ branches connect the same pair of nodes (i.e., the conductors are in parallel), artificial nodes are assigned to m-1 of of these branches. Each such conductor is thus divided into two conductors in series, and the conductances and emf are conveniently distributed to the parts.

Table I is an example of how the circuit in Fig. 1 can be tabulated. At the expense of additional storage space, the lines could also be numbered independently, making the treatment of lines in parallel somewhat simpler.

Method of Solution

An iterative procedure based on a theorem first proved by Maxwell[4] is used to obtain the solution of the d-c network as a set of potentials at the various junctions. An initial guess is provided by the program but intelligent guesses can be substituted if desired. Usually the time required to prepare a set of intelligent guesses is not justified because the difference in machine computing time is only a few minutes. The solution is predetermined whenever the potentials of two or more junctions of the network are given and fixed. The power dissipated by the system is a minimum when the set of potentials corresponds to the exact solution. The program uses the method of steepest descent to approach the exact solution. A simplified flow diagram of this program is shown in Fig. 2.

The Institute for Advanced Study machine has an internal high-speed memory of 1,024 words and a slower magnetic drum memory of 2,048 words. Input and output are through punched cards. The program is designed for networks of less than 256 lines and 128 junctions. All the circuit information and computing program can be accommodated in the internal memory. For larger networks, the magnetic drum has to be used with great loss in speed. The solution is punched on cards and a tabulating machine is used to present it to the problem originator in printed form.

A simple Wheatstone bridge, a plane potential problem, and a temperature distribution problem were solved by this method. The computing time ranges from 2 minutes to less than 15 minutes.

Table I. Incidence Matrix Associated with Wheatstone Bridge of Fig. 3

Incident Junction Number	Number of Destination Node	Resistance
1	2	R_1
0	2	R_2
0	3	R_4
1	3	R_5
2	3	R_6

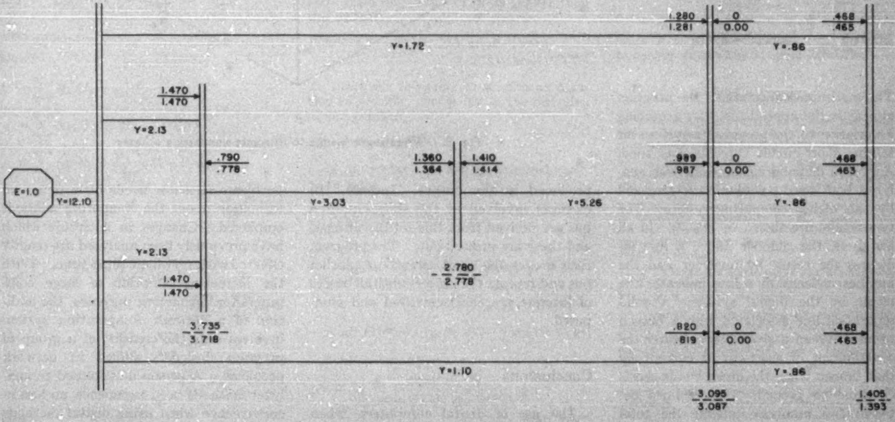


Fig. 1. Sample network, comparing analyzer and digital computer solutions

⟶ Line contribution
- - - - - - Bus total

Numbers above lines indicate currents obtained by analyzer; below lines, from digital computer; other numbers are conductances (Y)

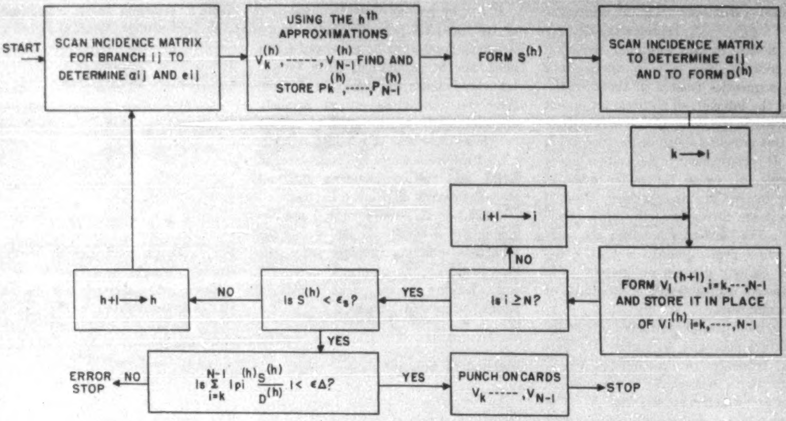

START
SCAN INCIDENCE MATRIX FOR BRANCH ij TO DETERMINE aij AND eij
USING THE h th APPROXIMATIONS $V_k^{(h)}, -----, V_{N-1}^{(h)}$ FIND AND STORE $P_k^{(h)}, ----, P_{N-1}^{(h)}$
FORM $S^{(h)}$
SCAN INCIDENCE MATRIX TO DETERMINE aij AND TO FORM $D^{(h)}$
k → i
i+1 → i
NO
FORM $V_i^{(h+1)}$, i=k,---,N-1 AND STORE IT IN PLACE OF $V_i^{(h)}$ i=k,----,N-1
Is i ≥ N?
YES
Is $S^{(h)} < \epsilon_s$?
NO
h+1 → h
YES
Is $\sum_{i=k}^{N-1} |p_i^{(h)} \frac{S^{(h)}}{D^{(h)}}| < \epsilon\Delta$?
NO
ERROR STOP
YES
PUNCH ON CARDS $V_k ----- , V_{N-1}$
STOP

Fig. 2. Flow diagram for finding potentials to minimize power loss by the method of steepest descent

N = number of junctions
V_i = potential at ith junction
a_{ij} = conductance in branch ij
$V_0, \ldots, V_{k-1}$ = known and fixed potentials

$$V_i^{(h+1)} = V_i^{(h)} - P_i^{(h)} \frac{S^{(h)}}{D^{(h)}}$$

$i = k, \ldots, N-1$

$$\frac{1}{2} P_i^{(h)} = \sum_{j=0}^{N-1} a_{ij}(V_i^{(h)} - V_j^{(h)} + e_{ij})$$

$i = k, \ldots, N-1$

$$S^{(h)} = \sum_{i=k}^{N-1} [P_i^{(h)}]^2$$

$$D^{(h)} = \sum_{i=k}^{N-1} \sum_{j=i+1}^{N-1} a_{ij}(P_i^{(h)} - P_j^{(h)})^2$$

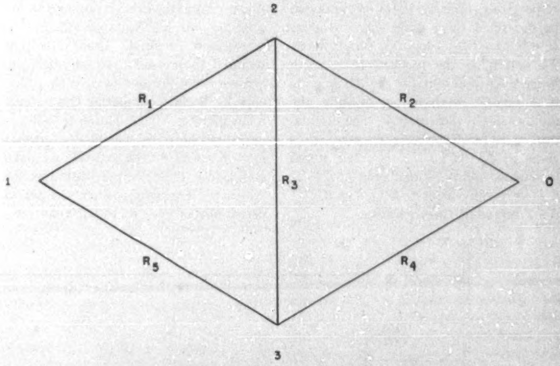

2
R1
R2
1
R3
0
R5
R4
3

Fig. 3. Wheatstone bridge to illustrate numbering scheme

The mathematical details of the program appear in the appendixes. By attaching a variation to the program, a system for solving short-circuit currents was tried. A typical distribution network was analyzed both by d-c network analyzer and by this digital computing system. The two results are shown in Fig. 3. In all instances, the number above a line indicates the result by analyzer and the number underneath a line indicates the result by the digital system. A solid separation line associated with a branch terminating on a given bus indicates the contribution of short-circuit current by that branch when the given bus is short-circuited to ground. Dotted lines between two numbers indicate the total current flowing into the short circuit. The generator junctions are fixed at unity voltage while the grounded bus is fixed at zero. The set of potentials of the solution is of no interest and is therefore not produced in the output. Instead, the currents involved at the short-circuited bus are derived from this set of voltages, and these are printed out. The program then moves the short circuit to another bus and repeats the process until all busses of interest are short-circuited and computed.

Conclusions

The use of digital computers, when developed into a system, is a practical way of solving complex network problems. It is demonstrated by the foregoing procedure that high-speed digital computers can be programmed such that the problem originator needs little or no information about the computing methods employed. Changes in a circuit which have previously been analyzed are readily effected with very little setup time. With the increasing adoption of large computers for accounting purposes, the addition of a network computation system involves only the creation of a group of program designers skilled in network problems. A person accustomed to analyzer methods need experience no loss in convenience when using digital methods instead. It is suggested that a floating-point computing machine with a larger internal memory be used because with a fixed-point machine, scaling difficulties seriously limit the accuracy.

Vol. Pt. 1 1956

Appendix I. Solution of Network Equations by Minimization of Power

Let $V_0, \ldots, V_k$, $k \geq 1$, be the known potentials, which are to remain unchanged throughout the entire calculation. Let $V_{k+1}, \ldots, V_{N-1}$ be the desired potential, for which intelligent or standard guesses are initially provided. Let I_{ij} be the correct current in the branch ij of a linear resistive network. $R_{ij} = \frac{1}{a_{ij}}$ is the resistance in that branch. Let V_i be the correct (yet unknown) potential at the ith node, including external emf's, so that $I_{ij} = a_{ij}(V_i - V_j)$. Next, let W_i be an approximation to V_i, and T_{ij} be the associated approximation to I_{ij}, so that $T_{ij} = a_{ij}(W_i - W_j)$ and $T_{ij} = I_{ij} + \epsilon_{ij}$. Let

$$P^* = \sum_i \sum_j R_{ij} T_{ij}^2 = \sum_{i,j} a_{ij}(W_i - W_j)^2$$

summed over the relevant values of i and j.

Theorem: P^* is minimum if and only if $T_{ij} = I_{ij}$, all i, j.

Proof: When P^* is minimum, W_i satisfies:

$$0 = \frac{\partial P^*}{\partial W_i} = 2\sum_j a_{ij}(W_i - W_j) = \sum_j T_{ij},$$

all i.

Since I_{ij} and T_{ij} are currents, they must satisfy Kirchoff's first law (the equation of continuity),

$$\sum_j I_{ij} = 0, \quad \sum T_{ij} = 0 \text{ for all } i$$

so that

$$\sum_j \epsilon_{ij} = \sum_j (T_{ij} - I_{ij}) = \sum_j T_{ij} - \sum_j I_{ij} = 0$$

also.

Now,

$$\begin{aligned} P^* &= \sum_{i,j} R_{ij}(I_{ij} + \epsilon_{ij})^2 \\ &= \sum_{i,j} R_{ij}(I_{ij}^2 + 2I_{ij}\epsilon_{ij} + \epsilon_{ij}^2) \\ &= \sum_{i,j} I_{ij}^2 R_{ij} + 2\sum_{i,j}(V_i - V_j)\epsilon_{ij} + \\ &\qquad \sum_{i,j} R_{ij}\epsilon_{ij}^2 \end{aligned}$$

But

$$\begin{aligned} \sum_{i,j}(V_i - V_j)\epsilon_{ij} &= \sum_i V_i \sum_j \epsilon_{ij} - \\ &\qquad \sum_j V_j \sum_i \epsilon_{ij} = \\ &\sum_i V_i \sum_j \epsilon_{ij} + \\ &\qquad \sum_j V_j \sum_i \epsilon_{ji} = 0 \end{aligned}$$

Hence

$$P^* = \sum_{i,j}(I_{ij}^2 R_{ij} + R_{ij}\epsilon_{ij}^2)$$

$R_{ij} > 0$ whenever I_{ij} is defined. Consequently, P^* is least when and only when ϵ_{ij} in all the branches where $R_{ij} > 0$ is 0. The minimum power is the power actually dissipated in heat, $\sum_{i,j} I_{ij}^2 R_{ij}$, in terms of the correct currents.

Appendix II. Numerical Algorithm to Compute Solution

Although several efficient methods[5,6] of solving large systems of linear algebraic equations have been programmed for digital computers, it was felt desirable to use the method of steepest descent for purposes of simplicity in testing the performance of the proposed computing system. Whereas the former methods obtain the "exact" solutions in a finite number of steps by manipulating the matrix elements, the method of steepest descent is an iterative procedure, which in this case has the advantage of not requiring the storage of all the matrix elements. Since the method of steepest descent, although well known, is not too readily available in the engineering literature, it is of interest to present a derivation (based on a lecture by H. H. Goldstine[7]) and a description of its use here.

P, regarded as a function of $V_k, \ldots, V_{N-1}$, this $(N-k)$-tuplet being henceforth abbreviated as (a vector) $\mathbf{V}$, is a continuous, bounded, and positive definite function of these variables. These prerequisites permit the determination of $\mathbf{V}$ so as to make P minimum by the method of steepest descent.

Let $\mathbf{V}^0$ be an initial guess, and $\Delta\mathbf{V}$ a correction to it. It is desired to find $\Delta\mathbf{V}$ so as to minimize $P(\mathbf{V}^0 + \Delta\mathbf{V})$, or equivalently, $P(\mathbf{V}^0 + \Delta\mathbf{V}) - P(\mathbf{V}^0)$. Expand $P(\mathbf{V}^0 + \Delta\mathbf{V})$ in a Taylor series about $\mathbf{V}^0$

$$P(\mathbf{V}^0 + \Delta\mathbf{V}) = \sum_{m=0}^{\infty} \frac{1}{m!}(\Delta\mathbf{V} \cdot \text{grad})^m P(\mathbf{V}^0)$$

and, neglecting terms of order higher than the second,

$$P(\mathbf{V}^0 + \Delta\mathbf{V}) - P(\mathbf{V}^0) \doteq \Delta\mathbf{V} \text{ grad } P(\mathbf{V}^0) + 1/2(\Delta\mathbf{V} \cdot \text{grad})^2 P(\mathbf{V}^0)$$

To make the foregoing expression as negative as possible, choose $\Delta\mathbf{V}$ in the direction of the gradient,

$$\Delta\mathbf{V} = -t \text{ grad } P(\mathbf{V}^0)$$

t is a scalar to be determined so as to make

$$\frac{d}{dt}[P(\mathbf{V}^0 + \Delta\mathbf{V}) - P(\mathbf{V}^0)] = 0 =$$

$$\frac{d}{dt}\left\{-t \text{ grad}^2 P(\mathbf{V}^0) + \frac{t^2}{2}(\text{grad } P(\mathbf{V}^0) \cdot \text{grad})^2 P(\mathbf{V}^0)\right\} =$$

$$-\text{grad}^2 P(\mathbf{V}^0) + t\,(\text{grad } P(\mathbf{V}^0) \cdot \text{grad})^2 P(\mathbf{V}^0)$$

Solving for $\Delta\mathbf{V}$,

$$\Delta\mathbf{V} = -\left[\frac{\text{grad}^2 P(\mathbf{V}^0)}{(\text{grad } P(\mathbf{V}^0) \text{ grad})^2 P(\mathbf{V}^0)}\right] \text{grad } P(\mathbf{V}^0)$$

and the ith component of $\Delta\mathbf{V}^h$, $\Delta V_i^h \doteq -P_i^h S^h / D^h$. Because the Taylor series was truncated after two terms, the foregoing expression is only a first approximation to $\Delta\mathbf{V}$. It will be called $\Delta\mathbf{V}^0$. A second approximation may be obtained by repeating the entire calculation using $\mathbf{V}^1$ in place of $\mathbf{V}^0$; where $\mathbf{V}^1 = \mathbf{V}^0 + \Delta\mathbf{V}^0$.

The general inductive step is

$$V_i^{h+1} = V_i^h + \Delta V_i^h$$

The foregoing derivation for the general case for any P satisfying the stated conditions may be simplified if P represents the power in an electric network.

Introducing the abbreviations

$$P_i^h \equiv \frac{\partial P}{\partial V_i} = 2\sum_{j=0}^{N-1} a_{ij}(V_i^h - V_j^h + e_{ij}) - \qquad i = k+1, \ldots N-1,\ h = 0, 1, 2 \ldots$$

$$S^h \equiv \sum_{i=k+1}^{N-1} (P_i^h)^2$$

$$D^h \equiv \sum_{i,j} \frac{\partial P}{\partial V_i} \frac{\partial P}{\partial V_j} \frac{\partial^2 P}{\partial V_i \partial V_j} =$$

$$2\sum_{i=k+1}^{N-1}\left[\sum_{j=i+1}^{N-1} a_{ij}(P_i^h - P_j^h)^2 + \sum_{j=0}^{k} a_{ij}(P_i^h)^2\right]$$

$$t^h = -\frac{S^h}{D^h}$$

Then,

$$\Delta\mathbf{V}^h = -\frac{S^h}{D^h} \text{ grad } P(\mathbf{V})\big|_{\mathbf{V} = \mathbf{V}^h}$$

To see that under the afore-mentioned hypothesis, the successive approximations, $\Delta\mathbf{V}^0$, $\Delta\mathbf{V}^1$, ... do converge to $\Delta\mathbf{V}$, substitute for $\Delta\mathbf{V}^h$

$$\begin{aligned} P(\mathbf{V}^{h+1}) - P(\mathbf{V}^h) &= -\frac{S^h}{D^h} \text{grad}^2 P(\mathbf{V}^h) + \\ &\quad 1/2\left(-\frac{S^h}{D^h}(\text{grad}P \cdot \text{grad})^2 P(\mathbf{V}^h)\right) = \\ &\quad -\frac{S^h}{D^h} S^h + 1/2\frac{(S^h)^2}{D^h} = -\frac{(S^h)^2}{2D^h} \end{aligned}$$

Since P is positive definite, $D>0$ and the foregoing expression is nonpositive.

References

1. Machine Computation of Power Network Performance, Discussions, **L. A. Dunstan.** *AIEE Transactions*, vol. 66, 1947, pp. 621–24.

2. On Electric Networks, **Wei-Liang Chow.** *Journal*, Chinese Mathematical Society, Shanghai, China, vol. II, no. 2, 1940, pp. 321–39.

3. Topologic and Algebraic Considerations in Network Syntheses, **R. M. Foster.** *Proceedings*, Symposium Modern Network Syntheses, Brooklyn Polytechnic Institute, Brooklyn, N. Y., 1953, pp. 8–18.

4. A Treatise on Electricity and Magnetism, **J. C. Maxwell.** Dover Publications, Inc., New York, N. Y., vol. I, pp. 407–08.

5. Numerical Computation of the Characteristic Values of a Real Symmetric Matrix, **W. Givens.** Oak Ridge National Laboratory Report No. *1574*, Oak Ridge, Tenn., February 19, 1954.

6. Method of Conjugate Gradients for Solving Linear Systems, **E. Stiefel, M. R. Hestenes.** National Bureau of Standards Report No. *1659*, U. S. Department of Commerce, Washington, D. C., March 10, 1952.

7. Numerical Mathematical Methods, **H. H. Goldstine.** Lectures given at the Moore School, 1946. "Theory and Techniques for Design of Electronic Digital Computers," vol. I, pp. 7.3–7.6.

(*Discussion follows.*)

Discussion

G. W. Bills (North American Aviation, Inc., Downey, Calif.): To read such a clear and orderly paper as this is pleasing. There is little doubt that for most cases the solution of load flow and short-circuit problems can best be carried out by an iteration of nodal equations.

At the Bonneville Power Administration we used the steepest descent method initially in the solution of load flow problems, but Gordon Smith of the Administration finally perfected a Gauss-Seidel technique which has proved far simpler and converges very well. A similar technique was also employed by the Institute of Numerical Analysis at Los Angeles where the same problem was worked on.

It also appears probable that acceleration techniques for decreasing the number of iterations when converging to a solution will be perfected in the near future, although this may be unnecessary with such high-speed computers as the *IBM 704*.

Here at North American Aviation we hope to set up a general program on the *IBM 704*, which is to be in operation this year, for the solution of all types of power system problems. We feel certain that no network analyzer in existence can compete timewise with such a computer once a general program is achieved.

The short-circuit current problem is a good example of the use of a digital computer to eliminate such tedious tasks as recording short-circuit values from a d-c board. In the program contemplated here the effects of mutual couplings between lines is also to be included.

Have the authors attempted the application of network topology along with digital computers to system problems?

I feel that this paper is a definite contribution to those engaged in network analysis. It appears that these digital methods can easily be extended to more difficult network problems such as the synthesis of networks.

S. Y. Wong and **M. Kochen:** The authors have not seriously looked into system problems other than the one described. However, there is some reason to believe that the proposed digital computing system with some possible variation in the choice of the mathematical algorithm can be employed to handle more complex d-c or a-c networks, or nonlinear networks, than the ones discussed.

It is possible to use the program to solve physical problems other than electrical network analysis: any physical problem for which an electric analogue can be made may be solved by the digital computing system. This includes potential problems of a certain kind, such as the temperature distribution problem mentioned in the paper.

Galvanometer Efficiency as a Design Parameter

F. K. HARRIS
MEMBER AIEE

IN THE ordinary use of a galvanometer, its response to a small voltage in a closed circuit is almost always the point of interest; its action can only be completely understood if the galvanometer is considered as a circuit element. The response of a galvanometer in the detector branch of a bridge or potentiometer circuit informs the operator of the direction and magnitude of the adjustment he must make to reduce the unbalance of the network. Reduced to its simplest terms, the galvanometer sees a certain resistance connected across its terminals in which a certain electromotive force (emf), the unbalanced voltage in the network, is present. The galvanometer acts in all respects like an electric motor which stalls after winding a spring (the galvanometer suspension). The useful work is represented by the energy stored in the suspensions of the deflected instrument. It is therefore proper to inquire what use the galvanometer makes of the power available at its terminals,[1] and what means can be used to make it utilize this power efficiently in its motor action.

Efficiency

If the voltage E which would appear across an open circuit in the galvanometer branch of an unbalanced network is considered as well as the circuit resistance R which is connected to the galvanometer terminals, the maximum power available to the galvanometer (considered as an energy receiver) is $E^2/(4R)$. If the time T is required for the galvanometer coil to move from its rest position to maximum deflection, the energy available to the galvanometer during the interval is $E^2T/(4R)$. The maximum deflection is logical for the observer to read whether or not he uses the technique of galvanometer reversal[2] to double effective sensitivity and to eliminate the effect of parasitic emf in the galvanometer itself. Energy storage in the suspension during the interval is $(^1/_2)\,U\theta^2$ where

U = the suspension stiffness
θ = the angular deflection

The motor efficiency can then be defined as the ratio of stored to available energy, or

$$\psi=\frac{(1/2)U\theta^2}{E^2T/(4R)}=2RU\theta^2/(E^2T) \qquad (1)$$

since the useful work of the stalling motor is done in winding its spring, i.e., is converted to potential energy in the twisted suspension.

T can be expressed as $T=T_0/(2\sqrt{1-\gamma^2})$, see Appendix, in terms of the free period T_0 of the galvanometer and the relative damping γ in the operating circuit, so that

$$\psi=4RU\theta^2\sqrt{1-\gamma^2}/(E^2T_0) \qquad (2)$$

Relative damping, defined explicitly in the Appendix, may be considered as the ratio of the existing damping to critical damping. Numerically it ranges from 0 to 1 for the underdamped galvanometer and is greater than 1 for the overdamped galvanometer. Alternatively, efficiency may be expressed in terms of the relative damping γ in the circuit, the open-circuit relative damping γ_0, and the ratio N of external critical damping resistance to galvanometer resistance. Thus it can be shown, see Appendix, that

$$\psi=\frac{4}{\pi}\sqrt{1-\gamma^2}(\gamma-\gamma_0)\left(1-\frac{\gamma-\gamma_0}{1-\gamma_0}\,\frac{1}{N+1}\right)\times\left(1+\epsilon^{-\frac{\pi\gamma}{\sqrt{1-\gamma^2}}}\right)^2 \qquad (3)$$

Efficiency is shown in Fig. 1 as a function of damping for various values of N, and for two values of open circuit damping. It is apparent from these families of

---

Paper 56-60, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 17, 1955; made available for printing November 18, 1955.

F. K. Harris is with the National Bureau of Standards, Washington, D. C.

The author is grateful for the wise counsel of the late Dr. Frank Wenner, who proposed this problem, and for the skill of James B. Griffin who constructed the galvanometer.

This work was supported in part by the Diamond Ordnance Fuze Laboratories and, in part, under a program of basic instrumentation in connection with the National Bureau of Standards, by the Office of Naval Research, the Office of Scientific Research of the Air Research and Development Command, and the Atomic Energy Commission.

curves that, for efficient operation, the ratio of external circuit resistance to galvanometer resistance should be high, and that the open-circuit relative damping should be as small as possible. The open-circuit damping should be kept small since energy utilized by the galvanometer for dissipative damping is converted to heat through eddy-current and windage losses during the motion of the coil, and is not available for storage as potential energy in the deflected system. It is not so obvious that, for high efficiency, the resistance of the circuit must be large compared to the resistance of the galvanometer. This apparent paradox is resolved when it is considered that the transient motional impedance of the galvanometer rather than its static resistance must be matched to the circuit for maximum energy transfer during the motion.

In practice it is usually not difficult to realize a design in which the ratio of circuit resistance to coil resistance is 10 or more at critical damping, and in which the open-circuit relative damping is no more than 0.01 or 0.02. Thus it may be expected that about 70 per cent of the energy available at the galvanometer terminals can be employed usefully during the interval in which the moving system goes from its rest position to maximum deflection. (The maximum energy utilization will be somewhat less if the resistance of the galvanometer suspensions is a significant part of the total resistance.)

Power and Energy Relations

It is interesting to observe the manner in which energy is accepted by the galvanometer during its motion, and the way in which the accepted power is apportioned among the various energy absorbing mechanisms.

The equation of motion of the moving-coil galvanometer is[3]

$$P\ddot{\theta}+[K+G^2/(R+R_g)]\dot{\theta}+U\theta=GE/(R+R_g) \tag{4}$$

where

P = the moment of inertia of the moving system
K = the coefficient of dissipative damping
G = the motor constant, i.e., the air-gap field multiplied by coil area turns

If the potential energy of the system at any instant is $(^1/_2)U\theta^2$, the potential power may be defined as the rate of energy storage in this form, the time derivative of potential energy. Similarly kinetic power may be defined as the time rate of change of kinetic energy, or

$$\frac{d}{dt}\left[(^1/_2)P\left(\frac{d\theta}{dt}\right)^2\right]$$

The damping power is $K(d\theta/dt)^2$, since $K(d\theta/dt)$ is the dissipative damping torque. The sum of these three quantities (potential + kinetic + damping power) comprises the total motor power of the system, that is ei, the product of the instantaneous back emf of the galvanometer and the current in the circuit. There is one additional component of power, i^2R_g, the joule power. The total power accepted by the galvanometer is the sum of motor power and joule power at any instant. The back emf is given by the expression $e=G(d\theta/dt)$, and the current is $i=(E-e)/(R+R_g)$.

Galvanometer current and back emf are plotted against time in Fig. 2 for a particular galvanometer in which the operating relative damping is 0.707. (This particular value of damping (0.707, or $1/\sqrt{2}$) was used to reduce the labor of computing power and energy relations, and because a relative damping of approximately this value is favorable to galvanometer operation in many instances.) The open-circuit relative damping is 0.017, and the ratio of circuit resistance to coil resistance is about 10 to 1 at critical damping. Figs. 3 and 4 show the distribution of power and energy respectively for the same galvanometer. The co-ordinates of all these graphs have been normalized, and the deflection is maximum when $\omega t=\pi$. It should be noted that, during the interval when the potential power is more than the power available from the circuit, the excess is supplied from kinetic power since the kinetic energy of the system is decreasing at this time. It may also be noted that relatively little energy is transformed into heat while the galvanometer is being deflected. A much larger amount is stored as potential energy. Under steady-state conditions, energy is used only to supply heat.

Galvanometer Design Considerations

In a specific application, the circuit resistance may be known with which the galvanometer is to be used as well as the voltage of unbalance that must be observed. With these data, the expression for efficiency furnishes a convenient starting point in the design of a suitable galvanometer.

The expression $\psi E^2T_0/(4R2\sqrt{1-\gamma^2}) = (^1/_2)U\theta^2$ shows the energy storage that may be expected in a given suspension during the time required to attain maximum deflection. This may be written

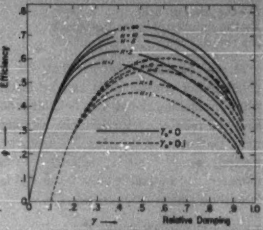


Fig. 1. Galvanometer efficiency as a function of damping

$$(\theta/E)^2=\psi T_0/(4RU\sqrt{1-\gamma^2}) \tag{5}$$

giving the expected voltage sensitivity in terms of the circuit resistance, the suspension stiffness, the free period of the galvanometer, and the relative damping at which it is to be operated. Alternatively, this expression may be used to calculate the combination of free period and suspension stiffness which will achieve a required sensitivity in a given circuit at a prescribed damping.

Since $T_0 = 2\pi\sqrt{P/U}$, sensitivity may also be expressed as a function of moment of inertia, suspension stiffness, and damping, in the relation

$$(\theta/E)^2=\pi\psi\sqrt{PU}/(2RU^2\sqrt{1-\gamma^2}) \tag{6}$$

If the physical size and the magnet structure of the instrument are specified, approximate values can be calculated for the moment of inertia P of the moving system and the stiffness U of the suspensions. Thus the maximum voltage sensitivity that can be achieved in a galvanometer design is limited by the length and cross section of the available suspension material as well as by the circuit resistance and the galvanometer period.

To complete the design the number of turns in the coil and the air-gap field must be specified. The motor constant of the galvanometer can be derived from the expression

$$R+R_g=G^2/[2\sqrt{PU}(\gamma-\gamma_0)] \tag{7}$$

Here it may be considered that the motor constant of the instrument is being fitted to the given circuit resistance for the desired value of operating relative damping. If the coil resistance can be kept low compared to circuit resistance (desirable for high efficiency), and if the open-circuit damping is small.

$$G\approx\sqrt{2\gamma R\sqrt{PU}} \tag{8}$$

which is nearly enough for present purposes. Now $G = BnA$ where

B = the air-gap field
n = the number of turns in the coil
A = the coil area

It will be seen therefore that, for a given motor constant, R_g can be made small by making the air-gap field large so that the number of turns in the winding may be small. Of course it must be realized that in a strong field the galvanometer will be stable only if the coil is free from magnetic impurities.

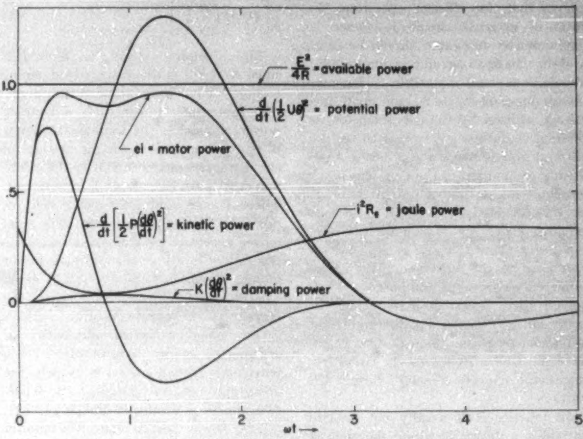


Fig. 3. Galvanometer power distribution during motion

DESIGN DETAILS

To illustrate the application of the foregoing equations, the design of a particular taut-suspension galvanometer will be considered in detail. The magnetic circuit is arranged oppositely to that generally used in galvanometer construction. The cylindrical core and shaped-pole pieces are of Alnico V. The core magnet and pole-piece structure are identical with those used in a particular Weston Sensitrol relay. The return magnetic circuit is a soft iron horseshoe to which the Alnico pole pieces are soldered. The air gap is 2 millimeters and has a field of 3,500 gausses. The coil is square, 11 millimeters on a side. No. 40 American Wire gauge copper wire was available, which had been especially prepared to be free from magnetic impurities, and was used in the construction. A collapsible winding form which accepts 15 turns of no. 40 wire per layer was used. To preserve mechanical symmetry in the moving system, two complete layers (30 turns) were wound on the form. This construction permitted the moment of inertia of the moving system to be computed accurately in advance. Only a portion of this total winding was used actively. Connections to the suspensions were made at one end of the coil and at an appropriate tap point.

The magnet structure and coil size being fixed, the design may proceed in either of two directions. If a particular voltage sensitivity at a chosen damping in a given circuit were desired, equation 6 could be used to select the suspension needed. Finally, equation 8 (or equation 7) would give the number of coil turns needed to fit the galvanometer to the given circuit. Alternatively, if it were desired to use suspensions of a particular length and cross section, their stiffness would be determined, and the coil turns required for operation at a chosen damping in the given circuit could be computed from equation 7 or 8. The expected voltage sensitivity and period could then be computed for the completed design. In either instance, the design equations describe the performance of the galvanometer with good accuracy.

In the particular design under consideration, it was desired that the galvanometer have the highest possible voltage sensitivity in a 50-ohm circuit, within the limits imposed by the available physical structure, with an assigned relative damping of 0.7. To this end, the smallest available suspension material was chosen. This was a 0.7-mil round wire of a copper-silver alloy having high strength and a conductivity over 70 per cent that of copper. (The ultimate strength of this wire, 150,000 pounds per square inch, corresponds to a loading of about 25 grams on the suspension.) Upper and lower suspensions 3 centimeters in length could be accommodated in an available mounting frame, and the stiffness of the suspensions could be computed. At this point the free period of the galvanometer could also be predicted from the relation $T_0=2\pi\sqrt{P/U}$. The predicted period, 1.38 seconds, was quite close to the actual period, 1.4 seconds, observed on the completed galvanometer, which indicated that the computed values of moment of inertia and stiffness were good.

The chosen suspensions have a combined resistance of about 6 ohms and, since a coil resistance of 3 ohms could be expected, the galvanometer resistance is

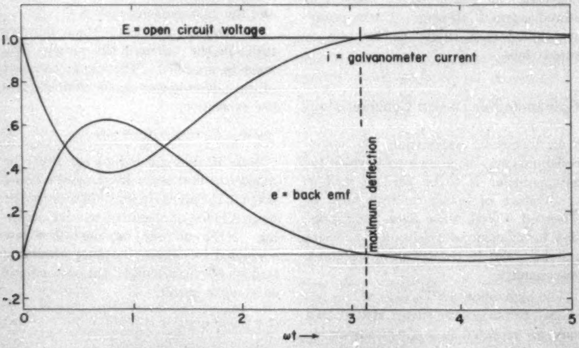


Fig. 2. Galvanometer current and back emf during motion

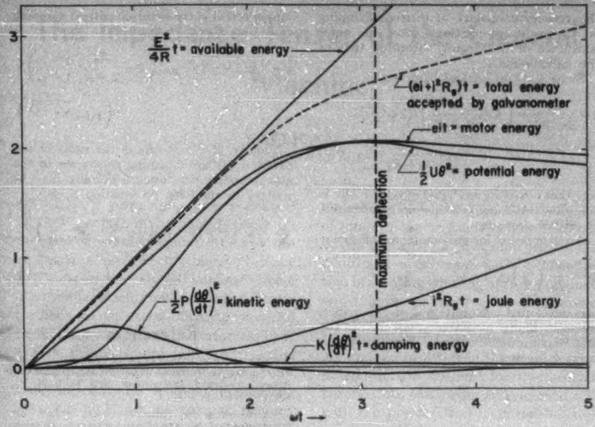


Fig. 4. Galvanometer energy distribution during motion

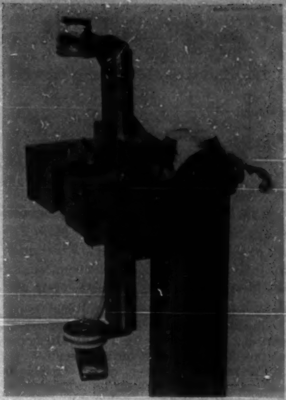


Fig. 6. Side view of galvanometer

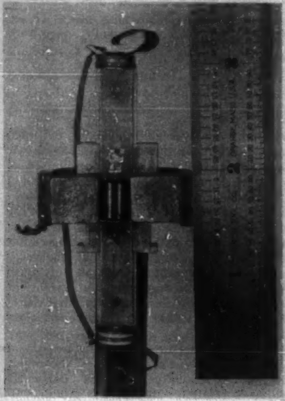


Fig. 5. Front view of galvanometer

9 ohms. Thus, in a 50-ohm circuit, the ratio of circuit to galvanometer resistance is only about 5, so that equation 7 rather than equation 8 was used to compute the motor constant of the galvanometer. A value of 0.725 times 10^5 was obtained as the desired motor constant, and this (in the available air-gap field of 3,500 gausses) set the value 17.1 as the number of turns needed to match the galvanometer to the operating circuit at the desired damping. Since the current enters and leaves the coil at opposite ends, the closest possible approximation to the desired winding is $17^1/_2$ turns. It is to be expected, therefore, that the actual damping would be slightly higher than the desired value.

The expected voltage sensitivity in the 50-ohm circuit remains to be computed. Because of the small cross section and high resistance of the suspensions, and because of the rather low operating-circuit resistance, the ratio of circuit to galvanometer resistance was much lower (5 to 1) than is desirable for best motor efficiency. Thus, under the existing conditions, an efficiency of only about 0.6 would be realized. This value was used in equation 6, and the predicted voltage sensitivity was 0.9 millimeter per microvolt (at 1 meter) in the 50-ohm circuit. Fig. 5 shows the front view and Fig. 6 the side view of a galvanometer.

Conclusion

The operating efficiency of a moving-coil galvanometer has been expressed as the ratio of stored energy to energy available at its terminals during its motion, and the power and energy utilization have been calculated in a typical example. Simple equations are given for adapting a galvanometer design to a particular set of operating conditions and for calculating the resulting voltage sensitivity. The validity of these equations has been demonstrated in a practical design of a galvanometer of high sensitivity and short period.

The use of a magnet design intermediate between the usual horseshoe type (with soft iron pole pieces and core) and the core-magnet structure[3] recently developed for indicating instruments provides a very compact structure and permits the realization of a strong air-gap field and, consequently, of efficient operation in a galvanometer having a length of 3 inches and a maximum cross section of about 1 square inch.

Appendix. Development of Equations

The equation of motion of the moving-coil galvanometer can be stated in the form[4]

$$P\ddot{\theta}+[K+G^2/(R+R_g)]\dot{\theta}+U\theta=GE/(R+R_g)$$

where

P = the moment of inertia of the moving system
K = the coefficient of dissipative damping
U = the stiffness of the suspension
G = the motor constant (air-gap field × coil-area turns)
R = the external circuit resistance
R_g = the galvanometer resistance (coil + suspension)
E = the open-circuit voltage at the galvanometer terminals

The solution of this equation, for underdamped motion, is

$$\theta=\theta_F\left[1-\frac{T_1}{T_0}\epsilon^{-\frac{At}{2P}}\sin\left(\frac{2\pi t}{T_1}+\tan^{-1}\frac{4\pi P}{AT_1}\right)\right]$$

where

$\theta_F=GE/[U(R+R_g)]$ = the steady-state deflection
$T_0=2\pi\sqrt{P/U}$ = the free period
$T_1=T_0/\sqrt{1-A^2/(4PU)}$ = the actual period in the operating circuit

$A=K+G^2/(R+R_g)=$ the coefficient of total damping (including both dissipative and conservative damping)

This solution may be written in the form

$$\theta=\theta_F\left[1-\frac{1}{\sqrt{1-\gamma^2}}\epsilon^{\frac{-2\pi\gamma t}{T_0}}\times \sin\left(\frac{2\pi t}{T_0}\sqrt{1-\gamma^2}+\sin^{-1}\sqrt{1-\gamma^2}\right)\right]$$

where

$\gamma=AT_0/(4\pi P)=A/(2\sqrt{PU})$ is the relative damping (the ratio of the actual coefficient to that required for critical damping)

The value of the first maximum deflection is $\theta_1=\theta_F\left(1+\epsilon^{\frac{-\pi\gamma}{\sqrt{1-\gamma^2}}}\right)$ and it occurs at a time $T=T_0/(2\sqrt{1-\gamma^2})$ after the voltage E is impressed on the circuit.

The relation between circuit resistance and relative damping can be expressed as $R+R_g=(R_c+R_g)(1-\gamma_0)/(\gamma-\gamma_0)$.

where

$R_c=$ the external circuit resistance (usually called *CDRX*) required for critical damping

$\gamma_0=K/(2\sqrt{PU})=$ the relative damping on open circuit

It may be noted from this relation that the conservative part of relative damping is inversely proportional to the total resistance of the circuit, including that of the galvanometer.

Since $\gamma=1$ at critical damping

$$1=\frac{A_c}{2\sqrt{PU}}=\left(K+\frac{G^2}{R_c+R_g}\right)/(2\sqrt{PU})$$
$$=\gamma_0+\frac{G^2}{R_c+R_g}/(2\sqrt{PU})$$

so that $R_c+R_g=G^2/[2(1-\gamma_0)\sqrt{PU}]$ is the total resistance (circuit+galvanometer) at critical damping.

An expression for galvanometer efficiency

$$\psi=\frac{4RU}{T_0}\left(\frac{\theta}{E}\right)^2\sqrt{1-\gamma^2}$$

was developed as equation 2 of this paper, efficiency being calculated at the time of the first maximum deflection. For this maximum deflection

$$\frac{\theta}{E}=\frac{G}{(R+R_g)U}\left(1+\epsilon^{\frac{-\pi\gamma}{\sqrt{1-\gamma^2}}}\right)$$

and

$$\psi=\frac{2RG^2\sqrt{1-\gamma^2}}{\pi(R+R_g)^2\sqrt{PU}}\left(1+\epsilon^{\frac{-\pi\gamma}{\sqrt{1-\gamma^2}}}\right)^2$$

If the galvanometer design parameters G, P, and U are eliminated, and the actual circuit resistance is replaced by its equivalent in terms of the external critical damping resistance R_c

$$\psi=\frac{4}{\pi}\sqrt{1-\gamma^2}(\gamma-\gamma_0)\left(1-\frac{R_g}{R_c+R_g}\frac{\gamma-\gamma_0}{1-\gamma_0}\right)\times \left(1+\epsilon^{\frac{-\pi\gamma}{\sqrt{1-\gamma^2}}}\right)^2$$

Finally, if $N=R_c/R_g$ is the ratio of the external critical damping resistance to the galvanometer resistance, efficiency may be written in the form of equation 3, as

$$\psi=\frac{4}{\pi}\sqrt{1-\gamma^2}(\gamma-\gamma_0)\left(1-\frac{\gamma-\gamma_0}{1-\gamma_0}\frac{1}{N+1}\right)\times \left(1+\epsilon^{\frac{-\pi\gamma}{\sqrt{1-\gamma^2}}}\right)^2$$

References

1. Deflection Constants of the D-C Galvanometer at Specified Damping, P. M. Andress. *Review of Scientific Instruments*, New York, N. Y., vol. 24, 1953, p. 172.

2. A Method of Reducing the Effect of Disturbances in the Galvanometer Branch of a Potentiometer Circuit, F. Wenner. *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 22, 1939, p. 425.

3. The Weston Core Magnet System, J. H. Miller. *Weston Engineering Notes*, Newark, N. J., vol. 4, Feb. 1949, p. 1.

4. Electrical Measurements (book), F. K. Harris. John Wiley and Sons, Inc., New York, N. Y., 1952, chap. 3.

◆

Discussion

H. N. Hayward (University of Illinois, Urbana, Ill.): The use of the word "efficiency" in describing the effectiveness of galvanometer response in this instance does not seem to be quite appropriate. The standard definition of "efficiency"[1] is, "The efficiency of a device with respect to a physical quantity which may be stored, transferred or transformed by the device is the ratio of the useful output of the quantity to its total input. Unless specifically stated otherwise, the term *efficiency* means efficiency with respect to power." There seems to be no question regarding the appropriate application of the first sentence of the definition. This comment refers particularly to the second sentence of the definition which, in effect, states that if other than power efficiency is referred to, that fact should be specifically stated. In this instance, energy efficiency rather than power efficiency is involved. There are already so many definitions of efficiency, each of which is unique for a given application, that it would appear highly desirable to add a qualifying or limiting adjective in this instance so that there can be no doubt about just what is meant. The primary objective in selecting a galvanometer for use in a bridge or potentiometer circuit for making null measurements is to be able to detect a sufficiently small amount of unbalance to establish that no significant unbalance actually exists. It is also because of this fact that the use of the single term "efficiency" as being appropriately descriptive is questioned. Although some thought has been given to the matter, no definite recommendation as to just what term might or should be used in place of "efficiency" has been selected. It is suggested that the matter be given further study.

The paper effectively points out that the proper basis for selecting a galvanometer is not the one that has been generally accepted and usually taken for granted. It should be of interest to have the characteristics of some of the commercially available laboratory galvanometers analyzed on the basis given in the paper, to determine the circuit characteristics under which they might be applied most effectively.

It is suggested that an effort be made to interest the galvanometer manufacturers in adding whatever information may be necessary to the galvanometer catalogue data now published to permit galvanometers to be selected in the future on the basis outlined in this excellent paper.

Reference

1. Definitions of Electrical Terms. *ASA C42-1941*, American Standards Association, New York, N. Y., 1941, paragraph 05.50.060, p. 60.

F. K. Harris: As Prof. Hayward points out, the term "efficiency" is used here in a very special sense, and a descriptive or qualifying adjective should be employed. To describe this quantity completely the modifiers would have to indicate that it is the ratio of energy stored to energy available from the external circuit during the time required by the galvanometer to reach maximum deflection. It would appear hopeless to try to find a combination of modifiers of reasonable length which would tell the complete story. Yet this "efficiency" is very useful as a parameter in the design equation since it permits the design to be fitted to a particular operating circuit condition. The term "dynamic efficiency" might be acceptable. Perhaps this term would convey the suggestion that a transient condition is being described, and that therefore an energy rather than a power relationship is involved.

With reference to the need for additional data in galvanometer catalogues, the circuit resistance associated with a particular value of relative damping is a function of the difference between the total relative damping and the open-circuit relative damping. This relation is $(R_a+R_g)/(R_c+R_g)=(1-\gamma_0)/(\gamma_a-\gamma_0)$, where R_a is the circuit resistance corresponding to the relative damping γ_a and R_c (the external critical damping resistance of the galvanometer) is the circuit resistance corresponding to critical damping. It would therefore be most helpful if galvanometer manufacturers would give values of open-circuit relative damping for their instruments or, where this constant is very small, would indicate approximately the maximum value it might be expected to have. The galvanometer resistance R_g, and the external critical damping resistance are usually available as catalogue information. There are also situations in which it would be helpful to know how R_g is divided between coil and suspension resistance.

Vol. 7
Pt. 1
1956

The Impedance Tensor of the General Machine

YAO-NAN YU
NONMEMBER AIEE

TWENTY years ago, Kron published the first part of his pioneer work on the application of tensors to the analysis of rotating electric machinery.[1] Since that time, many valuable studies[2-6] have been made which firmly establish tensor methods as an extraordinarily powerful method for analyzing the performance of electric machines and other complicated physical networks. With the aid of this tool, the engineer can solve intricate network problems in routine manner, thereby obtaining solutions directly.

This paper advances a study of the impedance tensors of general electric machines which are of salient-pole construction, and have m stator windings and n rotor windings. No restrictions on phase symmetry, such as an equal number of turns or an equal spacing of phase axes, are assumed.

The first part of the paper deals with the impedance tensors of a general machine with stationary winding axes; this is essentially a general commutating machine and includes induction machines of the squirrel-cage type as a special class. The derivation of the general impedance tensor is begun for a machine with only one stator winding and one rotor winding, and is then extended to the general case of m stator windings and n rotor windings. The paper then deals with the impedance tensor of a general machine with moving winding axes; this includes both synchronous machines and induction machines of slip-ring type as special classes. Since stationary winding axes can be considered as a special case of the general moving winding axes, the result of the second part of the paper naturally includes that of the first part.

In Appendix I the "back electromotive force (emf) principle" is introduced. With the aid of the physical concept underlying this principle, the coefficient of rotation can be easily determined. In Appendix II the physical interpretation of additional speed emf due to introduction of rotating winding axes is examined, and a simple method for the calculation of additional speed impedance is derived. In Appendix III, equations 38 and 46 are derived in detail in the usual form of differential equations from cross-field theory. As this paper is concerned only with the presentation of principles, only one example of an application is given, in Appendix IV. In this a short-circuited single-phase alternator is analyzed.

Direct Impedance Tensor of General Machine with Stationary Winding Axes

The general machine is assumed to have one stator winding a and one rotor winding u, as shown in Fig. 1, both winding axes being stationary in space though winding u itself is rotating with rotor speed $p\theta$. When stationary 2-reaction axes d and q are chosen as reference axes, with d leading q by 90 degrees (which implies that the clockwise rotating conductors meet d first and q 90 degrees later) the component impedances of the machine can be expressed as in the following matrix

$$Z=\begin{array}{c|cc|cc|}
 & d_a & q_a & d_u & q_u \\
\hline
d_a & L_{da}p & 0 & M_{dudu}p & 0 \\
q_a & 0 & L_{qa}p & 0 & M_{quqa}p \\
\hline
d_u & M_{dadu}p & M_{qadu}p\theta & L_{du}p & M_{qudu}p\theta \\
q_u & -M_{daqu}p\theta & M_{qaqu}p & -M_{duqu}p\theta & L_{qu}p \\
\hline
\end{array} \qquad (1)$$

where

$$L_{da}=N_a'N_aP_d\sin^2\theta_a$$
$$L_{qa}=N_a'N_aP_q\cos^2\theta_a$$
$$M_{dadu}=N_a'N_uP_d\sin\theta_a\sin\theta_u$$
$$M_{qaqu}=N_a'N_uP_q\cos\theta_a\cos\theta_u$$
$$M_{daqu}=N_a'N_uP_d\sin\theta_a\cos\theta_u$$
$$M_{qadu}=N_a'N_uP_q\cos\theta_a\sin\theta_u$$
$$M_{duda}=N_u'N_aP_d\sin\theta_u\sin\theta_a \qquad (2)$$
$$M_{quqa}=N_u'N_aP_q\cos\theta_u\cos\theta_a$$
$$L_{du}=N_u'N_uP_d\sin^2\theta_u$$
$$L_{qu}=N_u'N_uP_q\cos^2\theta_u$$
$$M_{duqu}=N_u'N_uP_d\sin\theta_u\cos\theta_u$$
$$M_{qudu}=N_u'N_uP_q\cos\theta_u\sin\theta_u$$

where N', N = effective number of turns of the exciting and the excited windings respectively, and P_d, P_q = permeance of the respective axes. A sinusoidal distribution of permeance in the air gap is assumed. The subscripts attached to the inductance refer successively to the exciting and the excited coils. The coefficient of rotation g of equation 1 is given in equation 3, as determined in Appendix I.

$$g=\begin{array}{c|cc|cc|}
 & d_a & q_a & d_u & q_u \\
\hline
d_a & & 0 & & 0 \\
q_a & 0 & & 0 & \\
\hline
d_u & & 1 & & 1 \\
q_u & -1 & & -1 & \\
\hline
\end{array} \qquad (3)$$

To find the direct self- and mutual-impedance tensors of a and u, a few transformations are necessary. The transformation matrix is of the following form

$$C=\begin{array}{c|c|c|}
 & i_a & i_u \\
\hline
i_{da} & 1 & 0 \\
i_{qa} & 1 & 0 \\
i_{du} & 0 & 1 \\
i_{qu} & 0 & 1 \\
\hline
\end{array} \qquad (4)$$

$$C_tZC=\begin{array}{c|c|c|}
 & a & u \\
\hline
a & (L_{da}+L_{qa})p & (M_{duda}+M_{quqa})p \\
u & (M_{dadu}+M_{qaqu})p-(M_{daqu}-M_{qadu})p\theta & (L_{du}+L_{qu})p-(M_{duqu}-M_{qudu})p\theta \\
\hline
\end{array} \qquad (5)$$

The result is equation 5, which shows that the impedance of a whole coil is the total sum of the self- and mutual-impedances of its d-, q- components. Therefore the same result can be obtained by summing equation 1 the various items in each block separated by heavy lines.

Next, the general machine with two rotor windings u and v, as shown in Fig. 2, is considered. With d and q chosen as reference axes as before, the component impedances of the machine can be expressed as

$$Z=\begin{array}{c|cc|cc|}
 & d_u & q_u & d_v & q_v \\
\hline
d_u & L_{du}p & M_{qudu}p\theta & M_{dvdu}p & M_{qvdu}p\theta \\
q_u & -M_{duqu}p\theta & L_{qu}p & -M_{dvqu}p\theta & M_{qvqu}p \\
\hline
d_v & M_{dudv}p & M_{qudv}p\theta & L_{dv}p & M_{qvdv}p\theta \\
q_v & -M_{duqv}p\theta & M_{quqv}p & -M_{dvqv}p\theta & L_{qv}p \\
\hline
\end{array} \qquad (6)$$

Paper 56-122, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30-February 3, 1956. Manuscript submitted February 14, 1955; made available for printing December 22, 1955.

YAO-NAN YU is with the National Taiwan University, Taipei, Taiwan, China.

The author wishes to express appreciation for the assistance given by his colleagues, Prof. C. L. Sheng, Hsu Chang, and Hwei-Piao Hsu, in the preparation of this paper.

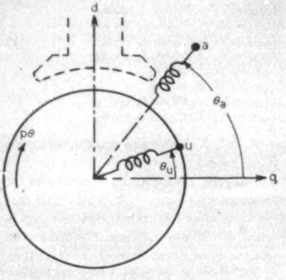


Fig. 1. One stator winding and one rotor winding of general electric machine

where

$$\begin{aligned}
L_{du}&=N_u'N_uP_d\sin^2\theta_u\\
L_{qu}&=N_u'N_uP_q\cos^2\theta_u\\
M_{duqu}&=N_u'N_uP_d\sin\theta_u\cos\theta_u\\
M_{qudu}&=N_u'N_uP_q\cos\theta_u\sin\theta_u\\
M_{dvdu}&=N_v'N_uP_d\sin\theta_v\sin\theta_u\\
M_{qvqu}&=N_v'N_uP_q\cos\theta_v\cos\theta_u\\
M_{dvqu}&=N_v'N_uP_d\sin\theta_v\cos\theta_u\\
M_{qvdu}&=N_v'N_uP_q\cos\theta_v\sin\theta_u\\
M_{dudv}&=N_u'N_vP_d\sin\theta_u\sin\theta_v\\
M_{quqv}&=N_u'N_vP_q\cos\theta_u\cos\theta_v\\
M_{duqv}&=N_u'N_vP_d\sin\theta_u\cos\theta_v\\
M_{qudv}&=N_u'N_vP_q\cos\theta_u\sin\theta_v\\
L_{dv}&=N_v'N_vP_d\sin^2\theta_v\\
L_{qv}&=N_v'N_vP_q\cos^2\theta_v\\
M_{dvqv}&=N_v'N_vP_d\sin\theta_v\cos\theta_v\\
M_{qvdv}&=N_v'N_vP_q\cos\theta_v\sin\theta_v
\end{aligned}\tag{7}$$

the coefficient of rotation being equation 8.

$$g=\begin{array}{c|c|c|c|c|}
 & d_u & q_u & d_v & q_v\\ \hline
d_u & & 1 & & 1\\ \hline
q_u & -1 & & -1 & \\ \hline
d_v & & 1 & & 1\\ \hline
q_v & -1 & & -1 & \\ \hline
\end{array}\tag{8}$$

After transformation, the direct self- and mutual impedances are as shown in equation 9

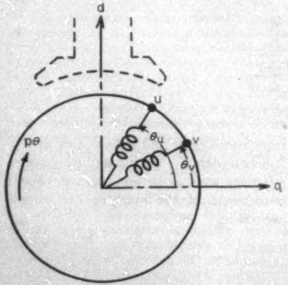


Fig. 2. Two rotor windings of general electric machine

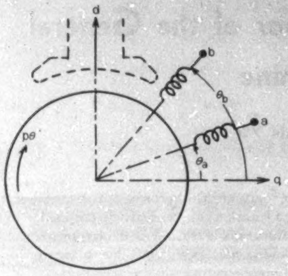


Fig. 3. Two stator windings of general electric machine

$$C_tZC=\begin{array}{c|c|c|}
 & u & v\\ \hline
u & (L_{du}+L_{qu})p-(M_{duqu}-M_{qudu})p\theta & (M_{dvdu}+M_{qvqu})p-(M_{dvqu}-M_{qvdu})p\theta\\ \hline
v & (M_{dudv}+M_{quqv})p-(M_{duqv}-M_{qudv})p\theta & (L_{dv}+L_{qv})p-(M_{dvqv}-M_{qvdv})p\theta\\ \hline
\end{array}\tag{9}$$

the transformation matrix being

$$C=\begin{array}{c|c|c|}
 & i_u & i_v\\ \hline
i_{du} & 1 & 0\\ \hline
i_{qu} & 1 & 0\\ \hline
i_{dv} & 0 & 1\\ \hline
i_{qv} & 0 & 1\\ \hline
\end{array}\tag{10}$$

Next, the general machine with two stator windings a and b, as shown in Fig. 3, is considered. The component impedances of this machine are the following

$$Z=\begin{array}{c|c|c|c|c|}
 & d_a & q_a & d_b & q_b\\ \hline
d_a & L_{da}p & 0 & M_{dbda}p & 0\\ \hline
q_a & 0 & L_{qa}p & 0 & M_{qbqa}p\\ \hline
d_b & M_{dadb}p & 0 & L_{db}p & 0\\ \hline
q_b & 0 & M_{qaqb}p & 0 & L_{qb}p\\ \hline
\end{array}\tag{11}$$

where

$$\begin{aligned}
L_{da}&=N_a'N_aP_d\sin^2\theta_a\\
L_{qa}&=N_a'N_aP_q\cos^2\theta_a\\
M_{dadb}&=N_a'N_bP_d\sin\theta_a\sin\theta_b\\
M_{qaqb}&=N_a'N_bP_q\cos\theta_a\cos\theta_b\\
M_{dbda}&=N_b'N_aP_d\sin\theta_b\sin\theta_a\\
M_{qbqa}&=N_b'N_aP_q\cos\theta_b\cos\theta_a\\
L_{db}&=N_b'N_bP_d\sin^2\theta_b\\
L_{qb}&=N_b'N_bP_q\cos^2\theta_b
\end{aligned}\tag{12}$$

The direct impedance tensor is given by

$$C_tZC=\begin{array}{c|c|c|}
 & a & b\\ \hline
a & (L_{da}+L_{qa})p & (M_{dbda}+M_{qbqa})p\\ \hline
b & (M_{dadb}+M_{qaqb})p & (L_{db}+L_{qb})p\\ \hline
\end{array}\tag{13}$$

the transformation matrix being

$$C=\begin{array}{c|c|c|}
 & i_a & i_b\\ \hline
i_{da} & 1 & 0\\ \hline
i_{db} & 1 & 0\\ \hline
i_{qa} & 0 & 1\\ \hline
i_{qb} & 0 & 1\\ \hline
\end{array}\tag{14}$$

Finally, the general machine in its broadest sense is considered. It has m stator windings and n rotor windings, all winding axes being stationary in space and arbitrarily spaced, as shown in Fig. 4. By referring to equations 5, 9, and 13, the direct-impedance tensor of this machine can be immediately written as the following:

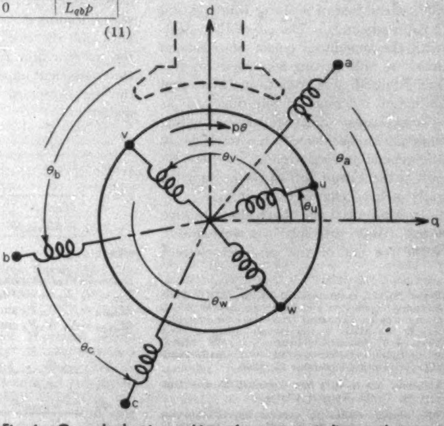


Fig. 4. General electric machine of m stator windings and n rotor windings, with 2-reaction axes d and q shown

$$C_tZC=\begin{array}{c|c|c|c|c|c|c|c|c|}
 & a & b & .. & c & u & v & .. & w \\ \hline
a & \begin{matrix} r_a+l_ap \\ +(L_{da}+L_{qa})p \end{matrix} & (M_{dbda}+M_{qbqa})p & .. & .. & (M_{duda}+M_{quqa})p & (M_{dvda}+M_{qvqa})p & .. & .. \\ \hline
b & (M_{dadb}+M_{qacb})p & \begin{matrix} r_b+l_bp \\ +(L_{db}+L_{qb})p \end{matrix} & .. & .. & (M_{dudb}+M_{quqb})p & (M_{dvdb}+M_{qvqb})p & .. & .. \\ \hline
.. & \dots & \dots & .. & .. & \dots & \dots & .. & .. \\ \hline
c & \dots & \dots & .. & .. & \dots & \dots & .. & .. \\ \hline
u & \begin{matrix} (M_{dadu}+M_{qaqu})p \\ -(M_{daqu}-M_{qadu})p\theta \end{matrix} & \begin{matrix} (M_{dbdu}+M_{qbqu})p \\ -(M_{dbqv}-M_{qbdu})p\theta \end{matrix} & .. & .. & \begin{matrix} r_u+l_up \\ +(L_{du}+L_{qu})p \\ -(M_{duqu}-M_{qudu})p\theta \end{matrix} & \begin{matrix} (M_{dvdu}+M_{qvqu})p \\ -(M_{dvqu}-M_{qvdu})p\theta \end{matrix} & .. & .. \\ \hline
v & \begin{matrix} (M_{dadv}+M_{qaqv})p \\ -(M_{daqv}-M_{qadv})p\theta \end{matrix} & \begin{matrix} (M_{dbdv}+M_{qbqv})p \\ -(M_{dbqv}-M_{qbdv})p\theta \end{matrix} & .. & .. & \begin{matrix} (M_{dudv}+M_{quqv})p \\ -(M_{duqv}-M_{qudv})p\theta \end{matrix} & \begin{matrix} r_v+l_vp \\ +(L_{dv}+L_{qv})p \\ -(M_{dvqv}-M_{qvdv})p\theta \end{matrix} & .. & .. \\ \hline
.. & \dots & \dots & .. & .. & \dots & \dots & .. & .. \\ \hline
w & \dots & \dots & .. & .. & \dots & \dots & .. & .. \\ \hline
\end{array} \qquad (15)$$

In equation 15, self-leakage impedances have been included. The various component inductances are

$$L_{da}+L_{qa}=N_a'N_a(P_d\sin^2\theta_a+P_q\cos^2\theta_a)$$
$$L_{db}+L_{qb}=N_b'N_b(P_d\sin^2\theta_b+P_q\cos^2\theta_b)$$
$$M_{dadb}+M_{qbqa}=N_b'N_a(P_d\sin\theta_a\sin\theta_b+P_q\cos\theta_a\cos\theta_b)$$
$$M_{dbda}+M_{qbqa}=N_b'N_a(P_d\sin\theta_b\sin\theta_a+P_q\cos\theta_b\cos\theta_a)$$
$$M_{dadu}+M_{qaqu}=N_a'N_u(P_d\sin\theta_a\sin\theta_u+P_q\cos\theta_a\cos\theta_u)$$
$$M_{daqu}-M_{qadu}=N_a'N_u(P_d\sin\theta_a\cos\theta_u-P_q\cos\theta_a\sin\theta_u)$$
$$M_{dbdv}+M_{qbqv}=N_b'N_v(P_d\sin\theta_b\sin\theta_v+P_q\cos\theta_b\cos\theta_v)$$
$$M_{dbqv}-M_{qbdv}=N_b'N_v(P_d\sin\theta_b\cos\theta_v-P_q\cos\theta_b\sin\theta_v)$$
$$M_{dadv}+M_{qaqv}=N_a'N_v(P_d\sin\theta_a\sin\theta_v+P_q\cos\theta_a\cos\theta_v)$$
$$M_{daqv}-M_{qadv}=N_a'N_v(P_d\sin\theta_a\cos\theta_v-P_q\cos\theta_a\sin\theta_v)$$
$$M_{dbdu}+M_{qbqu}=N_b'N_u(P_d\sin\theta_b\sin\theta_u+P_q\cos\theta_b\cos\theta_u)$$
$$M_{dbqu}-M_{qbdu}=N_b'N_u(P_d\sin\theta_b\cos\theta_u-P_q\cos\theta_b\sin\theta_u)$$
$$M_{duda}+M_{quqa}=N_u'N_a(P_d\sin\theta_u\sin\theta_a+P_q\cos\theta_u\cos\theta_a)$$
$$M_{dvdb}+M_{qvqb}=N_v'N_b(P_d\sin\theta_v\sin\theta_b+P_q\cos\theta_v\cos\theta_b)$$
$$M_{dudb}+M_{quqb}=N_u'N_b(P_d\sin\theta_u\sin\theta_b+P_q\cos\theta_u\cos\theta_b)$$
$$M_{dvda}+M_{qvqa}=N_v'N_a(P_d\sin\theta_v\sin\theta_a+P_q\cos\theta_v\cos\theta_a)$$
$$L_{du}+L_{qu}=N_u'N_u(P_d\sin^2\theta_u+P_q\cos^2\theta_u)$$
$$M_{duqu}-M_{qudu}=N_u'N_u(P_d\sin\theta_u\cos\theta_u-P_q\cos\theta_u\sin\theta_u)$$
$$L_{dv}+L_{qv}=N_v'N_v(P_d\sin^2\theta_v+P_q\cos^2\theta_v)$$
$$M_{dvqv}-M_{qvdv}=N_v'N_v(P_d\sin\theta_v\cos\theta_v-P_q\cos\theta_v\sin\theta_v)$$
$$M_{dudv}+M_{quqv}=N_u'N_v(P_d\sin\theta_u\sin\theta_v+P_q\cos\theta_u\cos\theta_v)$$
$$M_{duqv}-M_{qudv}=N_u'N_v(P_d\sin\theta_u\cos\theta_v-P_q\cos\theta_u\sin\theta_v)$$
$$M_{dvdu}+M_{qvqu}=N_v'N_u(P_d\sin\theta_v\sin\theta_u+P_q\cos\theta_v\cos\theta_u)$$
$$M_{dvqu}-M_{qvdu}=N_v'N_u(P_d\sin\theta_v\cos\theta_u-P_q\cos\theta_v\sin\theta_u)$$
$$\dots \qquad (16)$$

All the θ-angles in the foregoing equations are constant. P_d and P_q are equal for the nonsalient pole machine.

In Kron's analysis of electric machines, two primitive types are assumed as starting point. The first one consists of four windings (ds, qs, dr, qr), two (ds, qs) on the stator, and the other two (dr, qr) on the rotor, each pair having one winding along a common axis d and the other along another common axis q, axes d and q being at 90 degrees in space and all winding axes and reference axes being stationary in space. This corresponds to a special case of the general commutating machine already discussed. In equation 15, let

$$\theta_a=90°,\ \theta_b=0,\ \theta_u=90°,\ \theta_v=0 \qquad (17)$$

and assume that the machine has four windings a, b, u, v only. Then the impedance tensor of this machine can be written as equation 18.

The windings a, b, u, v, equation 18, correspond, respectively, to the windings ds, qs, dr, qr of Kron's first primitive machine. Rearranging the order, we change equation 18 to equation 18(A), which checks with Kron's result,[7] given for comparison in equation 19.

$$C_tZC=\begin{array}{c|c|c|c|c|}
 & a & b & u & v \\ \hline
a & \begin{matrix} r_a+l_ap \\ +N_a'N_aP_dp \end{matrix} & 0 & N_u'N_aP_dp & 0 \\ \hline
b & 0 & \begin{matrix} r_b+l_bp \\ +N_b'N_bP_qp \end{matrix} & 0 & N_v'N_bP_qp \\ \hline
u & N_a'N_uP_dp & N_b'N_uP_qp\theta & \begin{matrix} r_u+l_up \\ +N_u'N_uP_dp \end{matrix} & N_v'N_uP_qp\theta \\ \hline
v & -N_a'N_vP_dp\theta & N_b'N_vP_qp & -N_u'N_vP_dp\theta & \begin{matrix} r_v+l_vp \\ +N_v'N_vP_qp \end{matrix} \\ \hline
\end{array} \qquad (18)$$

$$C_tZC=\begin{array}{c|c|c|c|c|}
 & a & u & v & b \\ \hline
a & \begin{matrix} r_a+l_ap \\ +N_a'N_aP_dp \end{matrix} & N_u'N_aP_dp & 0 & 0 \\ \hline
u & N_a'N_uP_dp & \begin{matrix} r_u+l_up \\ +N_u'N_uP_dp \end{matrix} & N_v'N_uP_qp\theta & N_b'N_uP_qp\theta \\ \hline
v & -N_a'N_vP_dp\theta & -N_u'N_vP_dp\theta & \begin{matrix} r_v+l_vp \\ +N_vN_bP_qp \end{matrix} & N_b'N_vP_qp \\ \hline
b & 0 & 0 & N_v'N_bP_qp & \begin{matrix} r_b+l_bp \\ +N_b'N_bP_qp \end{matrix} \\ \hline
\end{array} \qquad (18A)$$

$$C_tZC=\begin{array}{c|c|c|c|c|}
 & ds & dr & qr & qs \\ \hline
ds & r_{ds}+L_{ds}p & M_dp & 0 & 0 \\ \hline
dr & M_dp & r_r+L_{dr}p & L_{qr}'p\theta & M_q'p\theta \\ \hline
qr & -M_d'p\theta & -L_{dr}'p\theta & r_r+L_{qr}p & M_qp \\ \hline
qs & 0 & 0 & M_qp & r_{qs}+L_{sq}p \\ \hline
\end{array} \qquad (19)$$

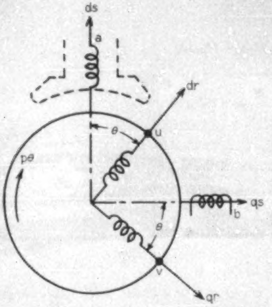


Fig. 5. Kron's second primitive machine

Direct-Impedance Tensor of General Machine with Moving Winding Axes

When moving winding axes are contemplated, additional speed emf are introduced which may be calculated by Christoffel matrix, as has been done by Kron.[8] However, as pointed out in Appendix II, these additional speed emf stem from the fact that the inductive coefficients which can be placed before p in equation 15 if constant, are no longer constants, and hence additional terms appear due to the necessity of differentiating the products involving the inductances. These additional speed reactances can be calculated by partially differentiating the exciting part of the inductive inductances and then replacing p with $p\theta$.

The terms of the impedance tensor of the general machine with moving winding axes can be derived from equation 15 with additional speed reactances added. A few typical cases are examined. Let the winding axes $u, v, \ldots, w$ of the general machine be assumed as rotating with the same speed of the rotor, while winding axes $a, b, \ldots, c$ remain stationary in space. First, $(M_{dadu}+M_{qaqu})p$ of equation 15 is considered. Since winding axis a is stationary in space and the exciting effect is not varying, there is no additional speed reactance in this case, hence

$$\frac{\partial}{\partial\theta_a}(M_{dadu}+M_{qaqu})p\theta_a=0 \tag{20A}$$

Next, $(L_{du}+L_{qu})p$ of equation 15 is considered. Since the winding axis u is moving in space and the exciting effect is varying, additional speed reactances appear in this case, and can be calculated as

$$\frac{\partial}{\partial\theta_u}(L_{du}+L_{qu})p\theta_u=N_u'N_u(P_d\cos\theta_u\sin\theta_u-P_q\sin\theta_u\cos\theta_u)p\theta_u \tag{20B}$$

Again, $(M_{dudv}+M_{quqv})$ of equation 15 is to be considered. Since the exciting effect is varying, additional speed reactances appear, which can be calculated as

$$\frac{\partial}{\partial\theta_u}(M_{dudv}+M_{quqv})p\theta_u=N_u'N_v(P_d\times\cos\theta_u\sin\theta_v-P_q\sin\theta_u\cos\theta_v)p\theta_u \tag{20C}$$

To obtain the terms of the impedance tensor of the general machine with moving axes, the additional speed reactances should be added to the corresponding terms of equation 15. For the cases studied

$$(M_{dadu}+M_{qaqu})p-(M_{daqu}-M_{qadu})p\theta_u=p\times N_a'N_u(P_d\sin\theta_a\sin\theta_u+P_q\cos\theta_a\cos\theta_u) \tag{21A}$$

$$(L_{du}+L_{qu})p-(M_{duqu}-M_{qudu})p\theta+\frac{\partial}{\partial\theta_u}(L_{du}+L_{qu})p\theta_u=p\times N_u'N_u\times(P_d\sin^2\theta_u+P_q\cos^2\theta_u) \tag{21B}$$

$$(M_{dudv}+M_{quqv})p-(M_{duqv}-M_{qudv})p\theta+\frac{\partial}{\partial\theta_u}(M_{dudv}+M_{quqv})p\theta_u=p\times N_u'N_v\times(P_d\sin\theta_u\sin\theta_v+P_q\cos\theta_u\cos\theta_v) \tag{21C}$$

where $p\theta=-p\theta_u$ in equations 21(A) and 21(B), and $p\theta=-p\theta_v$ in equation 21(C). Additively, $(L_{da}+L_{qa})p$ too can be written as $p(L_{da}+L_{qa})$, where $p\theta_a=0$.

In conclusion, the general expression for any term of the impedance tensor of the general machine with moving axes is $N_m'N_np(P_d\sin\theta_m\sin\theta_n+P_q\cos\theta_m\cos\theta_n)$.

Attention is called to the fact that the operator p operates not only on the inductance but also on the exciting current i_m. The angles θ may be constant or variable and may be different or equal. By referring to this general expression, the whole impedance tensor of the general machine with moving and/or stationary winding axes can be written immediately as follows:

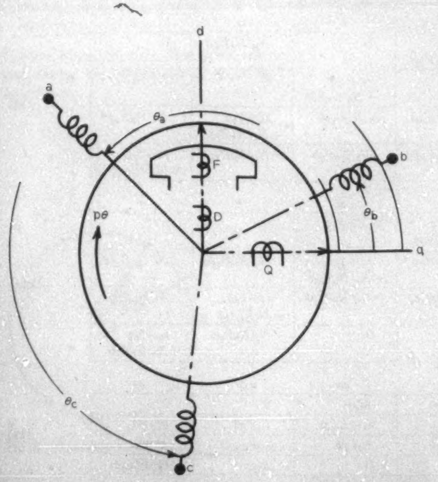


Fig. 6 (left). Three-phase 6-winding synchronous machine

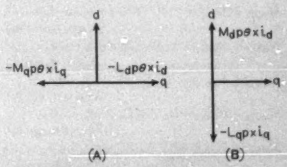


Fig. 7. Phase-angle relationship of the inductive and speed emf

A—For d-winding
B—For q-winding

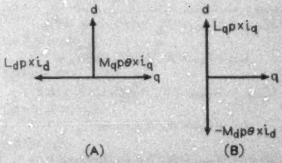


Fig. 8. Phase-angle relationship of applied inductive and speed voltages

A—For d-winding
B—For q-winding

$$
C_tZC=\begin{array}{c|c|c|c|c|c|c|}
 & a & b & c & u & v & w \\
\hline
a & r_a+p1_a+N_a'N_ap\times(P_d\sin^2\theta_a+P_q\cos^2\theta_a) & N_b'N_ap(P_d\sin\theta_b\sin\theta_a+P_q\cos\theta_b\cos\theta_a) & \cdot\cdot & N_u'N_ap(P_d\sin\theta_u\sin\theta_a+P_q\cos\theta_u\cos\theta_a) & N_v'N_ap(P_d\sin\theta_v\sin\theta_a+P_q\cos\theta_v\cos\theta_a) & \cdot\cdot \\
\hline
b & N_a'N_bp(P_d\sin\theta_a\sin\theta_b+P_q\cos\theta_a\cos\theta_b) & r_b+p1_b+N_b'N_bp\times(P_d\sin^2\theta_b+P_q\cos^2\theta_b) & \cdot\cdot & N_u'N_bp(P_d\sin\theta_u\sin\theta_b+P_q\cos\theta_u\cos\theta_b) & N_v'N_bp(P_d\sin\theta_v\sin\theta_b+P_q\cos\theta_v\cos\theta_b) & \cdot\cdot \\
\hline
 & \cdots & \cdots & \cdot\cdot & \cdots & \cdots & \cdot\cdot \\
\hline
u & N_a'N_up(P_d\sin\theta_a\sin\theta_u+P_q\cos\theta_a\cos\theta_u) & N_b'N_up(P_d\sin\theta_b\sin\theta_u+P_q\cos\theta_b\cos\theta_u) & \cdot\cdot & r_u+p1_u+N_u'N_up\times(P_d\sin^2\theta_u+P_q\cos^2\theta_u) & N_v'N_up(P_d\sin\theta_v\sin\theta_u+P_q\cos\theta_v\cos\theta_u) & \cdot\cdot \\
\hline
v & N_a'N_cp(P_d\sin\theta_a\sin\theta_v+P_q\cos\theta_a\cos\theta_v) & N_b'N_vp(P_d\sin\theta_b\sin\theta_v+P_q\cos\theta_b\cos\theta_v) & \cdot\cdot & N_u'N_vp(P_d\sin\theta_u\sin\theta_v+P_q\cos\theta_u\cos\theta_v) & r_v+p1_v+N_v'N_vp\times(P_d\sin^2\theta_v+P_q\cos^2\theta_v) & \cdot\cdot \\
\hline
w & \cdots & \cdots & \cdot\cdot & \cdots & \cdots & \cdot\cdot \\
\hline
\end{array} \qquad (22)
$$

In equation 22 self-leakage impedances have been included.

A few applications are given below as illustrative of use of the equations. First, the impedance tensor of Kron's second primitive machine may be derived directly from equation 22. This machine differs from Kron's first one in that all the rotor-winding axes are now rotating with the same speed as the rotor, while the stator-winding axes remain stationary in space.

Fig. 5 shows the second primitive machine, where no windings other than a, b, u, v are included, and

$$\theta_a=90^\circ,\ \theta_b=0,\ \theta_u=90^\circ-\theta,\ \theta_v=-\theta \qquad (23)$$

Substituting equation 23 into equation 22, the order rearranged, equation 22 becomes

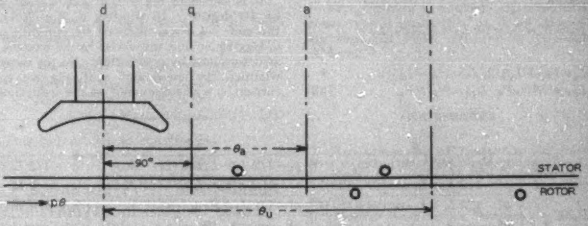


Fig. 9. Relative position between stator winding a and rotor winding u, with d-axis chosen as reference

$$
C_tZC=\begin{array}{c|c|c|c|c|}
 & a & u & v & b \\
\hline
a & r_a+1_ap+N_a'N_aP_dp & N_u'N_aP_d(\cos\theta p-\sin\theta p\theta) & -N_v'N_aP_d(\sin\theta p+\cos\theta p\theta) & 0 \\
\hline
u & N_a'N_uP_d\times(\cos\theta p-\sin\theta p\theta & r_a+1_up+N_u'N_u(P_d\cos^2\theta+P_q\sin^2\theta)p-N_u'N_u(P_d-P_q)\sin 2\theta p\theta & -N_v'N_u(P_d-P_q)\times[\sin\theta\cos\theta p+(\cos^2\theta-\sin^2\theta)p\theta] & N_b'N_uP_q\times(\sin\theta p+\cos\theta p\theta) \\
\hline
v & -N_a'N_vP_d\times(\sin\theta p+\cos\theta p\theta) & -N_u'N_v(P_d-P_q)\times[\sin\theta\cos\theta p+(\cos^2\theta-\sin^2\theta)p\theta] & r_v+1_vp+N_v'N_v(P_d\sin^2\theta+P_q\cos^2\theta)p+N_v'N_r(P_d-P_q)\sin 2\theta p\theta & N_b'N_vP_q\times(\cos\theta p-\sin\theta p\theta) \\
\hline
b & 0 & N_u'N_bP_q(\sin\theta p+\cos\theta p\theta) & N_v'N_bP_q(\cos\theta p-\sin\theta p\theta) & r_b+1_bp+N_b'N_bP_qp \\
\hline
\end{array} \qquad (24)
$$

which is identical with Kron's result[9] when a, u, v, b are replaced with ds, a, b, qs respectively, and the number of turns of windings u and v are assumed equal. For comparison, Kron's result is

$$
C_tZC=\begin{array}{c|c|c|c|c|}
 & ds & a & b & qs \\
\hline
ds & r_{ds}+L_{ds}p & M_d(\cos\theta p-\sin\theta p\theta) & -M_d(\sin\theta p+\cos\theta p\theta) & 0 \\
\hline
a & M_d(\cos\theta p-\sin\theta p\theta) & r_r(L_{dr}\cos^2\theta+L_{qr}\sin^2\theta)p+2(L_{qr}-L_{dr})\sin\theta\cos\theta p\theta & (L_{qr}-L_{dr})[\sin\theta\cos\theta p+(\cos^2\theta-\sin^2\theta)p\theta] & M_q(\sin\theta p+\cos\theta p\theta) \\
\hline
b & -M_d(\sin\theta p+\cos\theta p\theta) & (L_{qr}-L_{dr})[\sin\theta\cos\theta p+(\cos^2\theta-\sin^2\theta)p\theta] & r_r+(L_{dr}\sin^2\theta+L_{qr}\cos^2\theta)p+2(L_{dr}-L_{qr})\sin\theta\cos\theta p\theta & M_q(\cos\theta p-\sin\theta p\theta) \\
\hline
qs & 0 & M_q(\sin\theta p+\cos\theta p\theta) & M_q(\cos\theta p-\sin\theta p\theta) & r_{qs}+L_{qs}p \\
\hline
\end{array} \qquad (25)
$$

Next, the impedance tensor of 3-phase synchronous machine is derived. Fig. 6 shows such a synchronous machine, with

$$\theta_a=\theta+90^\circ,\ \theta_b=\theta_a-120^\circ,\ \theta_c=\theta_a+120^\circ,\ \theta_F=90^\circ,\ \theta_D=90^\circ,\ \theta_Q=0 \qquad (26)$$

Six windings are to be considered; namely, the symmetric 3-phase armature windings a, b, c on the stator, and the asymmetric field winding F, and amortisseur windings D and Q, on the rotor. F and D are along d-axis, while Q is along q-axis. It is assumed that the rotor was stationary while the stator was rotating with respect to the rotor.

Substituting equation 26 into equation 22, the latter becomes equation 27, in which

$$N_a=N_b=N_c=N,\ L_d=N'NP_d,\ L_q=N'NP_q \qquad (27)$$

$$L_0=(L_d+L_q)/2,\ L_2=(L_d-L_q)/2$$
$$L_{dF}=N'N_FP_d,\ L_{Qq}=N_Q'NP_q,\ \ldots \qquad (28)$$

$$C_tZC=\begin{array}{c|c|c|c|}
 & a & b & c \\ \hline
a & r_a+p1+p(L_0+L_2\cos 2\theta) & p[-1/2\,L_0+L_2\cos(2\theta-120)] & p[-1/2\,L_0+L_2\cos(2\theta+120)] \\ \hline
b & p[-1/2\,L_0+L_2\cos(2\theta-120)] & r_a+p1+p[L_0+L_2\cos(2\theta-240)] & p[-1/2\,L_0+L_2\cos 2\theta] \\ \hline
c & p[-1/2\,L_0+L_2\cos(2\theta+120)] & p[-1/2\,L_0+L_2\cos 2\theta] & r_a+p1+p[L_0+L_2\cos(2\theta+240)] \\ \hline
F & pL_{dF}\cos\theta & pL_{dF}\cos(\theta-120) & pL_{dF}\cos(\theta+120) \\ \hline
D & pL_{dD}\cos\theta & pL_{dD}\cos(\theta-120) & pL_{dD}\cos(\theta+120) \\ \hline
Q & -pL_{qQ}\sin\theta & -pL_{qQ}\sin(\theta-120) & -pL_{qQ}\sin(\theta+120) \\ \hline
\end{array}\ +$$

$$\begin{array}{c|c|c|c|}
 & F & D & Q \\ \hline
a & pL_{Fd}\cos\theta & pL_{Dd}\cos\theta & -pL_{Qq}\sin\theta \\ \hline
b & pL_{Fd}\cos(\theta-120) & pL_{Dd}\cos(\theta-120) & -pL_{Qq}\sin(\theta-120) \\ \hline
c & pL_{Fd}\cos(\theta+120) & pL_{Dd}\cos(\theta+120) & -pL_{Qq}\sin(\theta+120) \\ \hline
F & r_F+L_Fp & L_{DF}p & 0 \\ \hline
D & L_{FD}p & r_D+L_Dp & 0 \\ \hline
Q & 0 & 0 & r_Q+L_Qp \\ \hline
\end{array} \qquad (27)$$

Again, the impedance tensor of a single-phase synchronous machine can be derived. But this is only a special case of the 3-phase machine. If only one winding a on the stator and one winding u on the rotor are considered

$$\theta_a=\theta+90^\circ,\ \theta_u=\theta_F=90^\circ \qquad (29)$$

After substitution, equation 22 or equation 27 becomes

$$C_tZC=\begin{array}{c|c|c|}
 & a & u \\ \hline
a & r_a+p(L_0'+L_2\cos 2\theta) & pM\cos\theta \\ \hline
u & pM\cos\theta & r_u+pL_u \\ \hline
\end{array} \qquad (30)$$

where

$$L_0'=1+L_0,\ L_u=N_u'N_uP_d,\ M=N'N_uP_d \qquad (31)$$

With a similar procedure to that just demonstrated, the direct impedance tensor of any conceivable type of electric machine can be derived readily from equation 22 without laborious transformation.

Appendix I. Back-Emf Principle and Coefficient of Rotation

The back-emf principle[10] is advanced as follows: "When a machine is treated by cross-field theory, the speed emf of any rotating winding constitutes the back emf of the applied voltage of that winding, whether applied inductively, conductively, or simultaneously." This principle finds frequent applications. For the present discussion it serves to clarify confusions in the determination of coefficient or rotation, by interrelating the phase relations of speed and inductive emf.

It is shown in Fig. 7 that axis d leads q by 90 degrees. In winding d, Fig. 7(A), the emf $-L_dp\times i_d$, induced by current i_d, is lagging behind the d-axis by 90 degrees, as determined by convention. In the same winding, the speed emf $-M_qp\theta i_q$ due to current i_q is the back emf of the inductive emf $-L_dpi_d$, according to the back-emf principle. Therefore, the two emf are opposite in direction. By the same reasoning, in winding q, Fig. 7(B), the speed emf $M_dp\theta i_d$ and the inductive emf $-L_qpi_q$ are opposite in direction. Hence, the coefficient of rotation of the speed emf can be expressed as in equation 32.

It is an established convention and a fact of a-c circuit theory, that in a winding the applied inductive reactance voltage is opposite in phase to the induced inductive emf. By analogy, the speed reactance voltage applied is also opposite in phase to the speed emf induced. Hence, for the inductive and speed reactance voltages shown in Fig. 8, the coefficient of rotation, equation 33, is the negative of equation 32.

$$\begin{array}{c|c|c|}
 & d & q \\ \hline
d & & -1 \\ \hline
q & 1 & \\ \hline
\end{array} \qquad (32)$$

$$\begin{array}{c|c|c|}
 & d & q \\ \hline
d & & 1 \\ \hline
q & -1 & \\ \hline
\end{array} \qquad (33)$$

Appendix II. Additional Speed Emf due to Introduction of Rotating References Axes

The general equation of induced emf can be expressed as equation 34

$$e=-\frac{d\psi}{dt}=-\left(\frac{\partial\psi}{\partial t}+\frac{\partial\psi}{\partial\theta}\frac{d\theta}{dt}\right) \qquad (34)$$

where ψ is the total flux linkage of the coil, $-\partial\psi/\partial t$, the inductive emf, and $-\partial\psi/\partial\theta\, d\theta/dt$, the ordinary speed emf.

In the case of a coil with constant inductance, the inductive emf can be expressed as

$$-\partial\psi/\partial t=-\partial/\partial t(Li)=-L\,\partial i/\partial t \qquad (35)$$

But if the coil inductance is varying with space variation of excitation such as in the case of Kron's second primitive machine wherein the exciting coil on the stator excites the rotor windings in different degrees when the rotor axes assume different positions in space, or in the case of a salient-pole synchronous machine wherein the mutual inductance between the armature windings varies, due to variation of permeance in space, the inductive emf should be expressed as follows

$$-\frac{\partial\psi}{\partial t}=-\frac{\partial}{\partial t}(Li)=-\left(L\frac{\partial i}{\partial t}+\frac{\partial L}{\partial\theta_1}\frac{d\theta_1}{dt}i\right) \qquad (36)$$

The additional speed emf in this case corresponds to Christoffel's speed emf. The general speed of the rotating reference axis is represented by θ_1. In the case of holonomic system, $\theta=\theta_1$, the last item of equation 36 can be expressed as

$$e=-\left(\frac{\partial L}{\partial\theta}\right)p\theta i \qquad (37)$$

Therefore, the additional speed reactance can be calculated simply by differentiating partially those parts of the inductive reactance which vary with the space variation of excitation, and then replacing p by $p\theta$.

Variation of coil inductance in space due to the motion of the excited coil could also be introduced. But its effect on the induced emf has already been taken into account in the ordinary speed emf.

Appendix III. Derivation of Equation 1 from Cross-Field Theory

It should be pointed out that the starting point of this paper is the cross-field theory of asynchronous machines and the 2-reaction theory of synchronous machines like that of Kron. Fig. 1 is developed into Fig. 9, which shows a primitive machine of salient-pole construction having one stator winding *a* and one rotor winding *u*. Winding axis *a* displaces a constant angle θ_a from the stationary reference axis *d*, and winding axis *u* displaces a constant angle θ_u from *d*. Since cross-field theory is conceived, it is a necessity from the very beginning to treat the self- and mutual inductions of *a*, *u* windings in terms of their *d*, *q* components.

In the expression of inductances, effective turns are used, which have different meanings for exciting windings and for excited windings. For the excited winding of usual type, the chording factor f_c and winding distribution factor f_w besides the number of turns in series per winding should be taken into account. For the exciting winding, other factors should be included in the effective turns, such as $4/\pi$ to pick out the space fundamental; $1/k_s$, the reciprocal of saturation factor, to segregate the air gap ampere-turns; $2/\pi$ to average the sinusoidally variated permeance; and even Carter's coefficient k_c should be included in the effective turns, though the latter two factors $2/\pi$ and k_c may also be included in the permeance as an alternative.

When cross-field theory is utilized, both transformer emf and speed emf are induced. Now, let the *d*, *q* components of winding *a* be *da*, *qa* and those of winding *u* be *du*, *qu*. Transformer emf are induced in *du* by *da*, in *da* by *du*, in *qu* by *qa* and in *qa* by *qu*. Speed emf are induced in *du* by *qa* and in *qu* by *da*. Since *qu* is leading *du* by 90 degrees in space in the direction of rotation of the rotor, the speed emf in *qu* is leading the transformer emf in *du* by 90 degrees in time when both are excited by *da*. These emf can be written as

$$E_{dadu} = -j\sqrt{2}\pi f N_{du}\phi_{da} \tag{38}$$

$$E_{daqu} = \sqrt{2}\pi f N_{qu}\phi_{da} \tag{39}$$

This very fundamental relation offers another method in addition to that described in Appendix I for the determination of the positive and negative signs before the speed inductances of the "coefficient of rotation" of equation 1. Since reactance voltages are opposite in phase to induced emf, the minus sign is attached to M_{daqu} to express the correct time phase of the speed reactance voltage according to equation 39. The same sign is given to M_{duqu}. With similar reasoning, M_{qadu} and M_{qudu} are given positive signs.

Now the complete set of fundamental equation of voltages of the component windings for the cross-field theory analysis may be written as

$$\begin{aligned}
v_{da} &= L_{da}pi^a + M_{duda}pi^u \\
v_{qa} &= L_{qa}pi^a + M_{quqa}pi^u \\
v_{du} &= M_{dadu}pi^a + M_{qadu}i^a p\theta + L_{du}pi^u + \\
&\qquad M_{qudu}i^u p\theta \\
v_{qu} &= M_{qaqu}pi^a - M_{daqu}i^a p\theta + L_{qu}pi^u - \\
&\qquad M_{duqu}i_u p\theta
\end{aligned} \tag{40}$$

Equations 40 correspond to equation 1 where L_{da}, M_{duda}, etc., are given in equation 2.

Equations 40 may be added in pairs and reduced to

$$\begin{aligned}
v_a &= (L_{da}+L_{qa})pi^a + (M_{duda}+M_{quqa})pi^u \\
v_u &= [(M_{dadu}+M_{qaqu})p - (M_{daqu} - \\
&\qquad M_{qadu})p\theta]i^a + [(L_{du}+L_{qu})p - \\
&\qquad (M_{duqu}-M_{qudu})p\theta]i^u
\end{aligned} \tag{41}$$

This corresponds to equation 5.

Appendix IV. Short-Circuit Analysis of Single-Phase Alternator

Equation 30 in the usual form of differential equations becomes

$$\begin{aligned}
V_a &= r_a i^a + p[(L_0'+L_2\cos 2\theta)i^a] + \\
&\qquad p(M\cos\theta i^u) \\
V_u &= p(M\cos\theta i^a) + r_u i^u + p(L_u i^u)
\end{aligned} \tag{42}$$

This is the fundamental equation of a general 2-winding machine and may be applied to single-phase alternator.

To evidence the infinite harmonic reflections on both stationary and rotating members of the single-phase alternator as predicted by many authors, a short circuit at the armature terminals is analyzed. Since a short circuit at armature terminals is equivalent to applying suddenly a negative voltage $-v_a$ to annul completely the applied voltage v_a, the alternating components of short-circuit currents can be approximately solved by the "constant-linkage theorem" as follows:

$$\begin{aligned}
-v_a &= p[(L_0'+L_2\cos 2\theta)i^a] + \\
&\qquad p(M\cos\theta i^u) \\
0 &= p(M\cos\theta i^a) + p(L_u i^u)
\end{aligned} \tag{43}$$

After integration equation 43 becomes

$$\begin{aligned}
V[\cos\alpha - \cos(\theta+\alpha)] &= (L_0+L_2\cos 2\theta)i^a + \\
&\qquad M\cos\theta i^u \\
0 &= M\cos\theta i^a + L_u i^u
\end{aligned} \tag{44}$$

where

$$v_a = V\sin(\theta+\alpha) \tag{45}$$

whence

$$i^a = \frac{V}{\omega}\;\frac{\cos\alpha - \cos(\theta+\alpha)}{L_0+L_2\cos 2\theta - M^2\cos^2\theta/L_u}$$

$$i^u = \frac{MV}{\omega}\;\frac{\cos\theta[\cos(\theta+\alpha)-\cos\alpha]}{L_u(L_0+L_2\cos 2\theta) - M^2\cos^2\theta} \tag{46}$$

When this equation is developed, both even and odd harmonics will appear. In the classic work by Doherty and Nickle,[11] i^u was not considered when the constant linkage theorem is applied. If cylindrical rotor type is dealt with, L_2 in equation 45 becomes zero, which conforms to the result of Lyon.[12]

References

1. The Application of Tensors to the Analysis of Rotating Electric Machinery (book), G. Kron. *General Electric Review*, Schenectady, N. Y., a serial beginning in vol. 38, Apr. 1935, p. 181.

2. Tensor Analysis of Networks (book), G. Kron. John Wiley & Sons, Inc., New York, N. Y., 1939.

3. A Short Course in Tensor Analysis for Electrical Engineers (book), G. Kron. John Wiley & Sons, Inc., New York, N. Y., 1942.

4. Equivalent Circuits of Electrical Machinery (book), G. Kron. John Wiley & Sons, Inc., New York, N. Y., 1951.

5. Transient Analysis of Rotating Machines and Stationary Networks by Means of Rotating Reference Frames, Y. H. Ku. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 943–57.

6. Matrix and Electric Machinery (book), J. Takeuchi, G. Omoto. Denkigakuen, Tokyo, Japan, 1953.

7. See reference 3, p. 124, equation 16-10.

8. The Application of Tensors to the Analysis of Rotating Electrical Machinery (book), G. Kron. *General Electric Review*, Schenectady, N. Y., second edition, 1942, p. 62, equation 40.

9. See reference 8, p. 72, equation 6.

10. Alternating Current Machinery (book), Y. Yu. Cheng Chung Book Company, Taipei, China, 1955, p. 452.

11. Synchronous Machines—IV, R. E. Doherty, C. A. Nickle. *AIEE Transactions*, vol. 47, Feb. 1928, pp. 457–92.

12. Transient Analysis of Alternating Current Machinery (book), W. V. Lyon. John Wiley & Sons, Inc., New York, N. Y., 1954, p. 156.

Objectives and General Description of the Type-P1 Carrier System

R. C. BOYD
ASSOCIATE MEMBER AIEE

ALTHOUGH carrier has been used successfully to provide trunks in the Bell System for more than 35 years, it has not been economically feasible to apply carrier telephone techniques extensively to the rural telephone plant up to the present time.

The technical and economic problems associated with providing telephone service to subscribers in rural areas has long been one of the most difficult problems facing the telephone industry. The widely scattered locations of subscribers in rural areas have led to a large number of rural telephone routes with only a few subscriber lines per route. This has precluded the use of large cables on any one route as would be economically attractive in urban areas. The extensive use of carrier has not been feasible because the distances from the rural subscribers to the central office are in the 5- to 25-mile range in which carrier has not been generally economical in the past.

The two lines of attack which were taken on this problem were to reduce the cost of telephone plant through less expensive small cables and open-wire plant,[1] and to provide an economically attractive carrier system designed to meet the particular needs of rural carrier telephone service.

The *M1* system which has been in use since 1946 was the first successful Bell System effort to meet the need for more economical rural telephone service by carrier methods. It was developed primarily for application to commercial power lines but has also been used to a lesser extent on telephone lines. At present, there are roughly 10,000 subscribers served by *M* carrier channels.

Experience with the *M1* carrier system has indicated that carrier systems can be installed and maintained by exchange area forces with proper training and equipment. In addition, the background provided by the *M1* carrier indicated the direction for future development work on a rural subscriber carrier system.

The use of rural telephone lines indicated that in order to reduce the cost of line treatment it would be desirable to use frequencies lower than the 150- to 450-kc range used for the *M* carrier. In addition, lower frequencies would make it possible, from a crosstalk standpoint, to apply more than one system per rural line, which is often desirable in areas of rapid growth.

Inherent in the use of a rural subscriber system, such as the *M* carrier system, was the fact that a remote terminal must be placed on a pole or on a subscriber's premises. This created the need for low-power drains in the carrier equipment. The operating companies indicated the necessity of providing stand-by or primary battery power for the power supply of such a carrier terminal to insure reliable service in the face of commercial power failures.

In addition to the requirements for a rural subscriber system derived from the experience with the *M* carrier system, there were two other basic requirements: (1) to provide a grade of service equal to or better than that now available on long rural lines; and (2) to make available the several types of signaling required on Bell System multiparty lines served by manual or dial offices. However, the most important requirement of such a carrier system was that in addition to meeting the technical requirements imposed by the rural environment it also had to prove in *economically* over conventional physical telephone plant in the difficult 5- to 25-mile range.

The application of this carrier system background to the special technical and economic needs of rural telephone service has resulted in the development of the type *P1* rural subscriber carrier system. It is a fully transistorized system consisting of independent 2-way carrier channels applicable in increments of one to four at a time. Each channel is capable of providing single or multiparty (up to eight parties) exchange service to rural subscribers comparable to that provided by voice-frequency methods. The system will be applicable to existing and new lines utilizing combinations of fine-gauge exchange cable and copper or steel open-wire, repeaters to be used where necessary. Several systems can be used on a single pole line, the number depending on the open-wire line characteristics.

Development Background

To illustrate the development background of the *P1* system, the reasons behind the choices of a number of the parameters of the system are discussed in the following. Because of the stringent economic objectives of the system, cost was a paramount consideration in most of the items.

TRANSISTORS VERSUS VACUUM TUBES

Transistors were chosen for use in the *P1* system because they are low-voltage lower power devices as compared to the higher d-c power requirements of vacuum tubes suitable for transmission circuitry. Also, transistors are expected to be inherently longer life devices than vacuum tubes, thus contributing to reduced maintenance costs.

It was appreciated that the problems of transistor production and resulting component reliability were not completely resolved at the time the decision was made but it was felt that they showed sufficient promise to justify exploration of their possible use. This decision opened the way for use of primary battery power plants or those operated from commercial power with miniature storage battery reserve as described in another paper,[2] thus greatly reducing the cost of providing power at remote terminal and repeater locations.

LOW-VOLTAGE PROTECTION

With the use of transistors and the associated low-voltage power supplies, it became possible to use other low-voltage components in the carrier circuitry. This gave rise to the need for supplementary protection of such components from voltage surges on the line below those for which conventional carbon blocks afford protection. Voltage-limiting diodes offered a means of providing such protection. Development work was started to find and utilize diodes which could withstand voltages up to 1,000 volts.

FREQUENCY RANGE AND METHOD OF MODULATION

It is difficult to divorce the considerations leading to the choice of the carrier frequency range from those affecting the choice of modulation in the carrier system.

Paper 56-240, recommended by the AIEE Wire Communication Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing January 4, 1956.

R. C. BOYD is with Bell Telephone Laboratories, Inc., New York, N. Y.

The frequency range was blocked out by a number of factors.

Cost considerations required that the carrier frequencies be kept above the voice-frequency range. If carrier extended into the voice range, the voice-frequency circuit would be lost on a carrier pair. Then one of four carrier channels applied to that pair must be used to replace it. Thus, the addition of four carrier channels to a pair would yield a net gain of only three channels. This in turn would increase the net cost per gained channel. Filter costs determined how close to the voice-frequency band the carrier frequency range could be placed.

Crosstalk considerations restricted the carrier frequency range to below about 100 kc in order to reduce the cost of line treatment and rearrangement of pairs on existing rural lines. By using this frequency range it appeared possible to apply more than one carrier system to an open-wire line. The rapid rise in attenuation with frequency of steel wire used on rural lines also dictated that the range of carrier frequencies be kept low.

As a result of these two sets of considerations, development work on the *P1* carrier system was concentrated in the 8- to 100-kc range. At the time the choice was made, amplitude modulation of the carrier frequencies was chosen over other forms of modulation because of the simpler terminal circuitry and equipment. The use of amplitude modulation coupled with the compandors discussed in a later section was felt to compensate for possible transmission advantages which could be obtained by using angular modulation (frequency or phase) with a large modulation coefficient.

Cost was again a major factor in the choice between double-side-band and single-side-band amplitude modulation. Double side-band was chosen because of easier filter requirements. Although single-side-band modulation would give more efficient use of the frequency spectrum, it was ruled out because of the problems of obtaining necessary carrier power at the receiving terminal, and a possible high cost of filtering and circuitry to cut apart the carrier and the unwanted side-band.

Studies of growth on rural lines indicated that a system giving three or four channels on one pair of wires in addition to the physical circuit should be sufficient if several systems could be applied to a given open-wire line. It was, therefore, possible to space eight carriers at 12-kc intervals from 12 to 96 kc.

The unusually wide carrier spacing of 12 kc was adopted in order to minimize filter costs. Since the remote terminals are distributed along the line, it is not practical to use group modulation to accomplish filtering in the most efficient frequency range. Instead, filtering is done at line frequencies. This results in the use of rather large inductors and capacitors for some channels, thus placing a premium on the simple filter structures which wide spacing permits.

Fig. 1 shows the frequency plan for the *P1* carrier system. The staggered frequency arrangement was provided to improve crosstalk performance between systems, where necessary. The application of the three different frequency arrangements is discussed in detail in another paper.[3]

The use of frequencies spaced at 12-kc intervals for the *P1* system will create problems if applied to lines used by toll carrier systems which have 8-kc spacings. However, the cost advantages of the filters for 12-kc spaced carriers in general more than offset these potential problems.

System Levels and Carrier Line Loss

The carrier-frequency output power of the transistorized transmitting amplifier in the terminals was set at 6 decibels above 1 milliwatt. This was determined by the state of the transistor art, d-c power limitations at remote terminals, and co-ordination considerations. Because of the loss in the secondary protection circuitry, this became 4 decibels above 1 milliwatt at the carrier line terminals. Span pads were provided in the terminals and repeaters for use as required for level co-ordination.

This output level coupled with noise and crosstalk considerations indicated that a 30-decibel carrier line loss between the terminals would be possible. A survey of existing and planned Bell System rural telephone lines indicated that substantial amounts of entrance cable and open wire would be encountered in potential carrier channel layouts. Calculations of carrier frequency loss of these facilities indicated that carrier repeaters would be required between the terminals in order to serve the more remote subscribers. Fig. 2 shows an example of such data for one area studied.

Compandors

Compandors were incorporated in the *P1* system design because their added equipment cost was considered to be more than offset by their several advantages. The crosstalk and noise advantage provided by their use reduced the need for expensive line treatment to reduce crosstalk. In addition, it permitted lower receiver carrier levels to be used, thus increasing the permissible carrier line loss. Compandors also eased the requirements on terminal and repeater filters, thus reducing filter cost.

A number of different compandor configurations were studied including instantaneous and tone-controlled compandors before it was decided to use a simplified version of the compandor incorporated in the type *N* and *O* carrier systems.

Regulation

Channel regulation was deemed necessary to provide satisfactory transmission performance and to keep maintenance adjustments to a minimum. The regulation was designed to compensate for daily and seasonal carrier circuit net loss variations caused by changes in line attenuation with temperature. It would be desirable to have the terminal regulation range equal to the 30-decibel loss permitted between terminals to ease engineering layout considerations. However, cost considerations led to a smaller regulation range with span pads used as required by system layout to adjust the received carrier to within the regulation range.

Repeater regulation will be required as an option where variations in line loss exceed the terminal regulating range. It will usually be necessary on systems employing more than one repeater in order to control noise performance. Repeater regulation in the direction of transmission from central office to remote terminal will be controlled by the total carrier power of the one to four channels working in one system. In the opposite direction the repeaters will regulate on a carrier frequency pilot, because the channel carriers are not always present in that direction of transmission. The pilot frequencies are shown in Fig. 1. Span pads were also provided in the repeater to adjust the received carrier where necessary.

Signaling

The need to transmit subscriber signaling information over a carrier channel required the development of means of passing dialing and supervision signals toward the central office. It also required passing ringing information to the remote terminal for the types of multiparty ringing generally used in the Bell System. At that location standard 20-cycle ringing signals had to be recreated to ring the subscriber instruments. Carrier on-off signaling was chosen to transmit dialing and supervision signals because of the ease of implementing it and because of savings in d-c power drain of the remote terminal achieved by transmitting the carrier toward the central office only during the off-hook condition.

In-band frequencies were chosen to encode ringing information for transmission over the carrier channel because of the substantially lower cost of in-band filters as compared to those required for out-of-band transmission.

Summary of System Parameters

The system features evolved during the planning and development of the *P1* carrier system may be summarized as follows:

Transistors:

Diode low-voltage protection.

Frequency range:

9–99 kc.
Channel bandwidth 230 to 3,000 cycles at 5-decibel points.

Amplitude modulation—double side-band.
Three to four carriers per pair of conductors:

Staggered or normal arrangement.
Several systems per open-wire line.

System levels:

6 decibels above one milliwatt at transmitting amplifier output.
4 decibels above one milliwatt at carrier line terminals.

Carrier line loss:

30 decibels between terminals or repeaters.

Compandors:

Integral part of terminals.

Regulation:

15 decibels in terminal.
10-decibel option in repeaters.

Signaling:

Dialing and supervision by on-off carrier.
20-cycle ringing via in-band tones.

Discussions of considerations leading to design choices in areas of circuitry, equipment, and power are presented in other papers.[2,3]

Line Networks

The proposed application of *P1* terminals and repeaters to rural lines created the need for several types of line networks or filters. Low-pass filters were required to isolate bridged subscribers' instruments from the carrier line and to reduce their bridging loss to passing carrier frequencies. High-pass filters were required to isolate the voice-frequency extensions of different carrier channels from each other or from the physical circuit. Other networks were required at the junction of the cable and open-wire pairs. Small cross-arm-mounted networks were designed using the simplest possible configurations in keeping with the transmission and economic objectives of the *P1* system. Their use and equipment aspects are discussed further in other papers.[2,4]

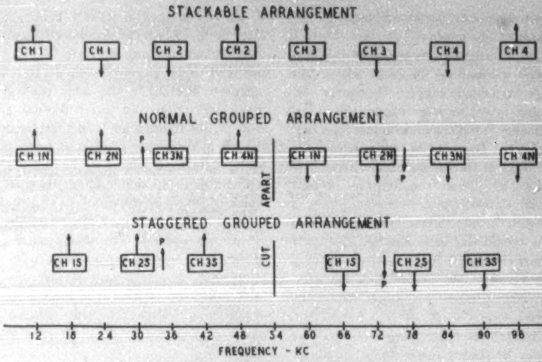


Fig. 1. Frequency allocation chart

P—Pilot frequency for repeater regulation

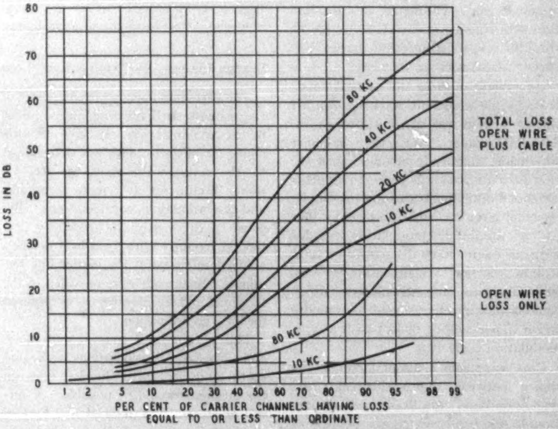


Fig. 2. Distribution of cable and open-wire losses for potential carrier channels

Maintenance

The prospect of the extensive use of the *P1* carrier system in the subscriber telephone plant led to the requirement that simple methods and test equipment be available for installing and maintaining the system. It was proposed that a single, compact test set be developed with options for battery or a-c operation, and that testing methods be devised to permit a craftsman with a minimum of training to isolate a carrier-line trouble to a unit of the system, such as a carrier terminal, and to adjust or replace that unit. Further isolation of trouble within a replaced unit could then be done most efficiently at a central test point. Alarms were not built into the *P1* system because it was felt that trouble reports should be received and analyzed at the exchange office test desk as is now the practice for physical rural lines.

Vol. 7
Pt. 1
1956

Present Status

Many of the circuits, components, and equipment arrangements incorporated in the *P1* carrier system had not been used or tested before. Therefore, experimental models of the carrier channel terminals, repeaters, line networks, and testing equipment were built and tested in the laboratory and in a representative rigorous rural environment at Americus, Ga.

The results of these tests were encouraging enough to justify the development of the *P1* carrier system for manufacture by the Western Electric Company. It is expected that the first channel terminals will be available to the Operating Companies in the Bell System in the second quarter, 1956. Repeaters will follow probably during the fourth quarter, 1956.

OTHER APPLICATIONS

The *P1* carrier system has been developed to provide an economical method of serving new rural telephone subscribers on new or existing rural telephone lines. It may also be used to give existing rural subscribers a higher grade of service, e.g., changing the service from 8-party semiselective to 4-party full selective ringing. In addition, the system will make it possible effectively to extend the ringing ranges of existing offices without adding other expensive switching equipment.

In addition to use as a rural subscriber system, the *P1* carrier system shows promise of providing a transmission facility for use on exchange area trunks. A new trunk signaling system is being developed for use in conjunction with the transmission portion of the *P1* carrier system, which is expected to be available for Bell System use early in 1957.

Conclusion

A study of the problem of providing economical rural subscriber telephone service indicated the need for an economical carrier system which could be operated on rural open-wire lines. The *P1* carrier system has been designed to meet that need by providing a flexible facility for adding a number of subscribers channels to an existing rural line with a minimum of line treatment.

References

1. TELEPHONE LINES FOR RURAL SUBSCRIBER SERVICE, Lester Hochgraf, R. G. Watling. *AIEE Transactions*, vol. 74, pt. I, May 1955, pp. 171–76.

2. POWER SUPPLIES FOR THE TYPE-P1 CARRIER SYSTEM, D. H. Smith. *Ibid.*, vol. 75, pt. I, May 1956, pp. 191–95.

3. CIRCUIT AND EQUIPMENT DESCRIPTIONS OF TYPE-P1 CARRIER SYSTEM, E. K. Eberhart, F. J. Hallenbeck, E. H. Perkins. *Ibid.*, pp. 195–204.

4. APPLICATION OF THE TYPE-P1 CARRIER SYSTEM, J. D. Howard, Jr. *Ibid.*, pp. 205–14.

Power Supplies for the Type-P1 Carrier System

D. H. SMITH
MEMBER AIEE

Synopsis: Low first cost and annual charges, good voltage regulation, high reliability, and low maintenance were major objectives guiding the choice of power supply design for the type-*P1* (rural) carrier system recently developed by Bell Telephone Laboratories. A study of conventional designs revealed that new techniques would be necessary to realize these objectives. This paper describes how the use of semiconductor devices, the employment of unconventional approaches to power supply packaging, and the development of a new type of lead-acid storage battery influenced design decisions to the end that the major objectives were realized.

EARLY in the development program[1] of the type-*P1* (rural) carrier system, engineering studies revealed that, for several reasons, the power supply problem would be difficult to solve. First, there was the large relative cost of supplying a small amount of power in the order of 1 watt. Second, there was the requirement for 1-per-cent or better voltage regulation under extreme environmental conditions. Third, there was the need for high reliability in locations remote from the telephone office and, concurrent with this, the necessity for reducing maintenance to a single service call per year. This paper describes how a power plant was developed to meet all four objectives.

Power requirements for the carrier system are quite low. The central office terminal draws about 75 ma (milliamperes) at 22 volts from the standard office power plant. A remote terminal averages about 50 ma at 22 volts, but the load varies from a quiescent current of about 30 ma to almost 1/2 ampere during the ringing interval. A repeater, serving up to four carrier channels,[2] requires about 150 ma at 22 volts. In addition, each equipment draws a few ma at 16 volts for bias. A remote power supply must operate satisfactorily in an ambient temperature of from −40 to +140 F (degrees Fahrenheit).

Commercial 60-cycle power is available in almost all rural areas in this country and the power plant described here is designed to operate from this source. Long secondary feeders are common, however, and some form of power reserve is necessary. When 60-cycle power is not available, a primary battery is furnished as an alternative to the a-c operated supply.

The High Cost of Low Power

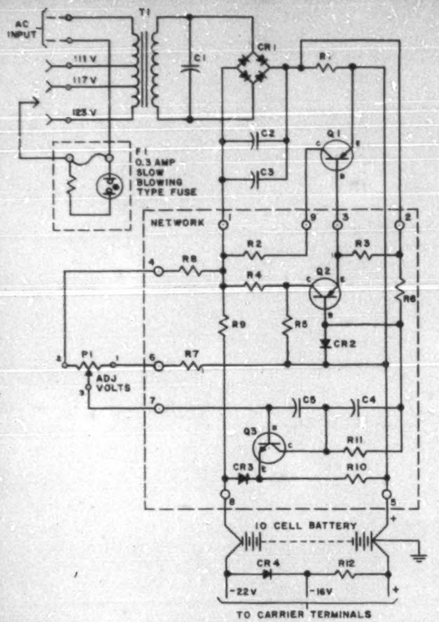
The cost of a telephone power plant generally decreases with decreasing capacity but, even so, the cost per watt-hour generally goes up as the capacity goes down. Experience has shown that a conventional design to deliver 1 or 2 watts costs nearly as much as one to deliver 10 to 20 watts. One yardstick often used is to relate the cost of the telephone power plant to that of the equipment which utilizes the power. This ratio can vary from 5 per cent for very large multiunit central offices to 40 or 50 per cent for very small installations. The cost objectives of the rural carrier system were such that past experience had to be bettered by at least 2 to 1. This was realized by using semiconductor devices for power rectification and control, the development of a new type of lead-acid storage battery to meet severe temperature conditions, and efficient equipment packaging.

For output ratings in the power range of 1 or 2 watts, prime sources tend to be even more expensive than conversion of commercial alternating current and pro-

---

Paper 56-243, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted November 22, 1955; made available for printing December 12, 1955.

D. H. SMITH is with the Bell Telephone Laboratories, Inc., New York, N. Y.



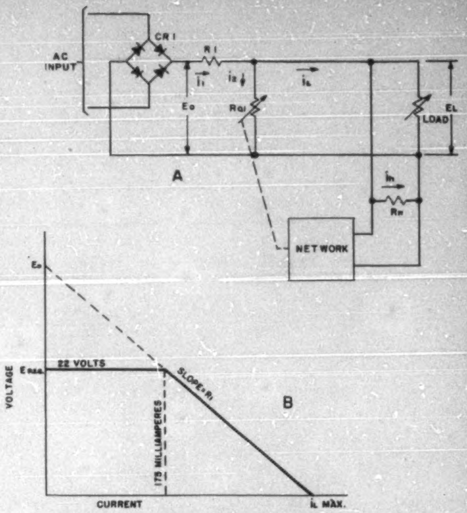


Fig. 1 (left). Power plant circuit

Fig. 2 (above). Simplified circuit

A—Equivalent circuit
B—Load characteristics

vision of a reserve storage battery. Storage batteries on a removal-for-charging basis, dry cells, wind generators, small internal combustion engines, and thermocouple generators were considered but, in each case, the economics favored an a-c operated rectifier with battery reserve.

The Power Plant Circuit

The rectifier, Fig. 1, uses semiconductor diodes *CR1* in a full-wave bridge circuit. In the first models, germanium diodes are furnished, but it is expected that these will be replaced by silicon in the near future. The output of *CR1* is fed to a transistor shunt regulator *Q1* through a ballast resistor *R1*. *Q1* is controlled by a transistor network (enclosed by the dashed lines). The operation of the network is to increase the emitter-base current of *Q1* when the battery voltage is too high and to decrease the current when the battery voltage is too low. An increase in the emitter-base current of *Q1* causes an increase in the collector-emitter current, an increased voltage drop in ballast resistor *R1*, and a subsequent reduction in battery voltage until the regulated value is reached. Transistor *Q1* is a p-n-p device having a maximum collector dissipation rating of about 2 watts at 140 F, when mounted on a heat-dissipating plate or chassis.

The heart of the control network is a varistor *CR3* which serves as a voltage reference standard because of its very stable voltage drop of about 6 volts over a wide range of temperature and current. The properties of this diode have been covered previously.[3] The emitter current of transistor *Q3* flows through the reference diode, and the base-to-emitter voltage is the difference between the voltage drop across terminals 2 and 3 of potentiometer *P1* and the reference voltage across *CR3*. With *P1* adjusted so that the base of *Q3* is slightly positive with respect to the emitter, an increase in the battery voltage will increase the base-to-emitter voltage and, thus, the base current of *Q3*. The collector current of *Q3* will then increase and, since part of it also flows through the emitter-base circuit of transistor *Q2*, through a similar action the collector current of *Q2* will increase. This in turn increases the collector current of transistor *Q1* which reduces the battery voltage, as covered previously. The total loop gain in the regulator circuit is about 35 decibels, and an over-all regulation accuracy of ±1 per cent is maintained over a temperature range of −40 to +140 F, input voltage variation of ±5 per cent, and load current range of 30 to 175 ma. The high loop gain also provides most of the ripple filtering necessary.

Above a 175-ma load, the load voltage decreases linearly to approximate a constant-current characteristic. Such protection is required when battery reserve is provided. Normally the battery "floats" across the load, drawing less than 1 ma or so at the regulated voltage. It undergoes a momentary discharge when load currents in excess of 175 ma are drawn, such as during the ringing interval. This energy is quickly restored, however. If a-c power fails, the battery supplies all the load current until it is fully discharged. Upon the return of a-c power after a prolonged outage, the battery tends to draw a very high current, far beyond the capabilities of the rectifier. The current-regulating action limits this recharge current to a safe value. This action is explained by reference to Fig. 2, the general equivalent circuit of the power plant.

In Fig. 2(A), *R1*, R_{Q1}, and R_N are respectively the regulating or ballast resistor, the shunt resistance of the emitter-collector circuit of transistor *Q1*,

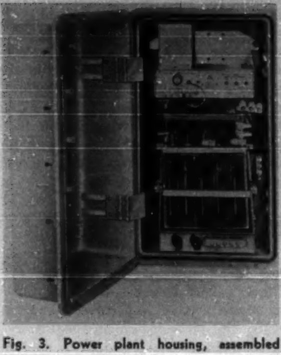


Fig. 3. Power plant housing, assembled

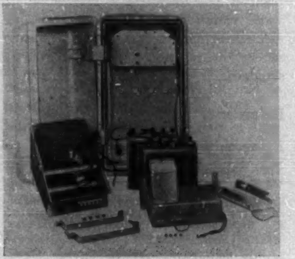


Fig. 4. Power plant housing, components removed

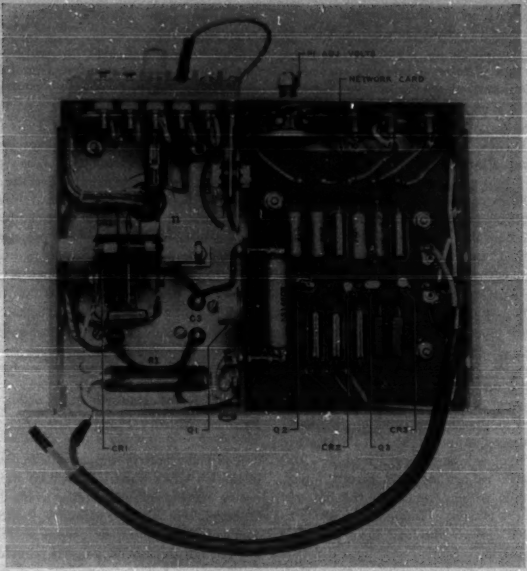


Fig. 5. Bottom view of rectifier chassis

and the equivalent resistance of the network components. The current drawn by the network i_n can be considered constant and part of the load current i_1. The load characteristic of the rectifier is represented by the heavy line in Fig. 2(B). At minimum load, R_{Q1} is very low and enough current i_2 is shunted around the load to drop the voltage from E_0 to the regulated value of 22 volts. As the load increases, R_{Q1} is made to increase, thus reducing I_2 and, in effect, transferring more and more of i_1 to the load. At full regulated load, determined by the intersection of the load line (slope $= R_1$) and the desired value of regulated voltage (22 volts), transistor *Q1* cuts off i_2. Further increase in load current causes a reduction in load voltage, as shown. *R1* can be varied within rather wide limits at the designer's discretion. An increase in *R1* will provide a sharper current cutoff but will raise the required transformer secondary voltage. The maximum power dissipation of *Q1* also will be increased.

The shunt-regulating circuit of Fig. 1 was chosen for this application because the current-limiting feature described in the foregoing is an inherent characteristic. A series-regulating circuit would be more efficient, but the additional circuitry required to provide current-limiting in a series regulator would add expense and complexity which could not be justified by the relatively slight power saving.

A final feature of interest in the rectifier is the provision of regulated base bias potential for the carrier terminal circuit. In Fig. 1, *CR4* and *R12* form a potentiometer across the 22-volt output. *CR4* is a 6-volt reference voltage diode similar to *CR3*, and acts as a second shunt regulator furnishing a potential always 6 volts more positive than the collector voltage.

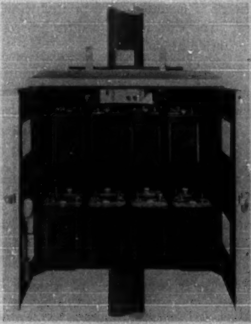


Fig. 6. Alternate power supply using primary batteries

The Battery

The battery weighs about 10 pounds and stores sufficient energy when fully charged to operate a carrier terminal about 4 days and a repeater about 2 days. In addition to its function of providing reserve during a-c power failure, the battery furnishes additional current required by the load during ringing intervals. These intervals are short and do not discharge the battery appreciably. This feature permits size reduction of the power transformer and permits conservative operation of the rectifying diodes and transistor *Q1*.

The electrolyte has a specific gravity of 1.300 or about 7 per cent higher than conventional telephone storage batteries. An excess of electrolyte is provided so that, when the battery plates are fully discharged, the electrolyte still has acid content to prevent freezing at lowest temperatures. Ordinarily, the higher acid content would increase local action and shorten the life of the battery, but a new type of battery plate grid, made of a lead-calcium alloy, inhibits local action so that a service life of about 10 years is expected.

Power Plant Packaging

The a-c operated power plant is housed in a pole-mounted cast-aluminum box similar to that provided for the carrier terminal, thus taking advantage of the lower cost with a single design. The box is about $23^{3}/_{4}$ inches tall, $12^{1}/_{2}$ inches wide, and $10^{5}/_{8}$ inches deep. Ready for operation, the plant weighs about 50 pounds. The power plant is shown in Fig. 3. At the top of the housing is the rectifier. The battery, consisting of two 5-cell plastic jars, produced in quantity for other applications, is arranged on "steps" so that the electrolyte level can be observed readily. Removable rubber stoppers permit filling the cells without removing the battery from the housing. Immediately beneath the battery are the load fuses and terminal strip.

In Fig. 4, the components of the power plant have been removed from the box and, in Fig. 5, a bottom view of the rectifier is given. The regulating control network is mounted on a printed-wiring card. The card is designed for low-cost mechanized assembly.

Reliability and Maintenance

The objectives of high reliability and low maintenance were weighed carefully against first cost throughout the development of the power plant. Transistors were operated at conservative levels for maximum life. Resistors and capacitors were chosen on the same basis. The storage battery was made slightly oversize to lengthen the maintenance interval; present estimates are that water need be added only at intervals of once a year or longer. In the network circuit, a reference diode having a slightly positive temperature coefficient is used to compensate for the small negative temperature coefficient of the emitter-base junction in transistor *Q3*. Thus, the necessity of a seasonal readjustment of the output voltage is eliminated.

Primary Battery Plant

A primary battery supply will be used in the few cases where commercial alternating current is not readily available. It consists of alkaline primary (nonrechargeable) batteries, as shown in Fig. 6. Enough battery capacity is provided to operate a remote terminal for about 2 years and a repeater for about 1 year. Specially designed to operate in extreme temperatures, the battery uses a caustic alkali as an electrolyte, a zinc negative electrode, and a carbon positive electrode, and is oxygen-depolarized. The batteries are shipped dry and water is added to liquify the dry electrolyte on the job. After the battery is fully discharged, a new battery is installed. The battery plant, installed and ready for service, measures $35^{1}/_{2}$ by $35^{1}/_{8}$ by 14 inches, and weighs about 350 pounds. It is more expensive, both in first cost and annual charges, than the a-c operated plant.

Conclusion

The a-c operated power plant described has been developed for the type *P1* (rural) carrier system. When operated at 50 ma average load current, and 22 volts, the plant maintains the battery at full charge and provides a reserve against failure of primary power. The plant is pole-mounted, is insensitive to wide variations in line voltage and ambient temperature, and will have a long service life consistent with Bell System standards.

References

1. OBJECTIVES AND GENERAL DESCRIPTION OF THE TYPE-P1 CARRIER SYSTEM, **R. C. Boyd.** *AIEE Transactions*, vol. 75, pt. I, May 1956, pp. 188–91.

2. CIRCUIT AND EQUIPMENT DESCRIPTIONS OF THE TYPE-P1 CARRIER SYSTEM, **E. K. Eberhart, F. J. Hallenbeck, E. H. Perkins.** *Ibid.*, pp. 195–204.

3. THE SUITABILITY OF THE SILICON ALLOY JUNCTION DIODE AS A REFERENCE STANDARD IN REGULATED METALLIC RECTIFIER CIRCUITS, **D. H. Smith.** *Ibid.*, vol. 74, pt. I, Jan. 1955, pp. 645–51.

◆

Discussion

Charles F. Crocker (General Electric Company, Lynn, Mass.): This paper is a very good presentation of what was done but I would like to see the reasons for doing it this way. Was any special selection necessary to get transistors of suitable characteristics for this unit? Is there any desirable range of α for *Q1*, *Q2*, and *Q3*? Because the circuit is direct-coupled throughout, the characteristics of *Q3* would seem to be critical. What I_{co} limits give satisfactory operating conditions for this stage? I would also like to see detailed characteristics for the power transistor *Q1*, particularly collector current, voltage rating, and power gain. It is gratifying to see the application of semiconductor devices to a product that must be reliable. This power supply uses several types of units with large savings in power and space and apparently an increase in reliable life over conventional equipment.

There is still a lack of acceptance of the newer types of these products by some sections of industry. I should be interested in seeing the data on the several transistor types and the junction diode types that led to their adoption on a basis of increased reliability. There are probably many people, such as myself, who are very concerned about reliability of electronic equipment but are not able to carry on any significant life testing. Mr. Smith, being associated with an organization that has done so much fundamental work on transistors, and who obviously knows what reliability of equipment means, may be able to give us some very valuable data.

There are many ways that such data could be presented, but I would like to suggest a few that would be of much interest to me. These are service failure rates in the following classes: 1. amount of derating from rated values of currents and voltages; 2. initial parameters such as α and I_{co}; 3. time in service; and 4. ambient operating conditions.

D. H. Smith: Mr. Crocker's comments relate to one of the most interesting design problems in many circuits using semiconductors, namely, that of the interdependence of circuit functions and device characteristics. Arrangements that appear to be straightforward in the notebook stage often become almost unrecognizable as the designer adds functions to overcome such variations as transistor current gain, transistor leakage current, and rectifier forward drop. Unfortunately, the treatment of this subject is not within the scope of the present paper.

During the development of the regulator under discussion, however, we were fortunate that semiconductor device development had already reached the point where manufacturing and operational variations in the parameters were well characterized. Circuit analysis, based on the range of current gain and collector saturation current for standard Western Electric transistors, indicated satisfactory circuit performance under specified environmental, load current, and line voltage conditions. The general statement can be made that, in production, special selection should not be required for any semiconductor components, nor should special transistor designs be necessary for the regulator circuit described.

The reliability question cannot, of course, be completely resolved until considerable operating experience is obtained. During the field trial of the *P1* carrier system, very good reports were received after initial circuit shakedown. Some data show, for example, that, of 90 semiconductor diodes employed in lines as lightning protectors, only one gave evidence after 2 years of operation of adverse drift in characteristics. Our observations of four regulated rectifiers of the type described in my paper, installed and operating in the Georgia field trial for about the same period of time, revealed no failures caused by transistors or diodes after initial shakedown.

Detailed life data in the form suggested by Mr. Crocker would be welcomed by circuit designers, and manufacturers should be encouraged to furnish life data on the device data sheets. Until this becomes a general practice, it appears that the design engineer must continue to rely on the integrity of the device manufacturer, together with judgment obtained from assiduous development experience.

Circuit and Equipment Descriptions of Type-P1 Carrier System

E. K. EBERHART
NONMEMBER AIEE

F. J. HALLENBECK
ASSOCIATE MEMBER AIEE

E. H. PERKINS
NONMEMBER AIEE

THE broad objectives and application principles for the *P1* carrier system opened a new field of use for carrier equipment. However, to exploit this field, several new requirements had to be met. These demanded that the system be of substantially lower cost than any before, that it use only a small fraction of the power, yet maintain the quality associated with a carrier system.

These objectives have been met. Carrier circuitry has been devised to use transistors which consume only a fraction of the power required by vacuum tubes. Plug-in equipment concepts have been worked out so that the manufacturing and handling of the apparatus is greatly improved, and new and better component parts have been developed to fulfill circuit and equipment needs.

This system is generally similar to the recently designed type-*N* carrier system. It utilizes double-side-band transmission which is appreciably simpler and cheaper than single-side-band transmission. It has compandors to liberalize crosstalk coupling requirements and to improve the signal-to-noise ratio. It uses the stackable arrangement of channels, with each channel a complete entity, but with the ability to stack channels in frequency to form a complete system.

This system, however, has important differences from previous systems. Whereas previous systems have been for use between central offices where both terminals were identical, this system is for use between a central office and a subscriber. This use requires terminals with somewhat similar transmission circuits but completely different signaling circuits. The signaling is further complicated by the variety of types of service offered to rural subscribers. Thus, the terminals for this system are specifically designed to the needs of the respective ends and are designated as the central office terminal and the remote terminal.

A detailed analysis of this system readily divides into the aspects of transmission, equipment, and component features. The system transmission may be further divided into terminal and repeater considerations.

Terminal Transmission

The transmission block schematic for the terminals is shown in Fig. 1. The voice circuit at either end of the system connects to the carrier circuit through a secondary protection circuit and resistance hybrid circuit. The protection offered by the usual carbon blocks was adequate for vacuum tubes but transistors require a much lower value. This additional protection is obtained by using the reverse-voltage breakdown characteristic of newly developed silicon diodes.

The hybrid circuit in turn connects to the compandor and signaling circuits. In the transmitting direction the carrier portion of the terminal consists of a modulator with associated carrier oscillator, a transmitting amplifier with level adjusting pads, and a transmitting filter. In the receiving direction there is a receiving filter, a level adjusting pad, and a receiving amplifier-demodulator. This carrier portion of the terminal is connected to the line through an impedance-matching transformer and secondary protection equipment.

The compandor though performing the same function as for previous systems employs different circuitry. The compressor uses a backward-acting shunt variolosser. The variolosser employs silicon diodes controlled by its rectified output signal, integrated over a syllabic period. As this variolosser operates at very low signal levels, a 2-stage feedback transistor amplifier is used to amplify the voice signals sufficiently to activate the modulator.

The 2-stage transistor-type amplifier is a basic building block of the system, and it is referred to as the *P*-terminal 2-stage amplifier. It is designed as a voice amplifier for the compressor and expandor and as a carrier amplifier for the transmitting and receiving amplifiers. A circuit drawing of this transmitting amplifier is shown in Fig. 2. It uses n-p-n type of transistors, coded *4B*, for the voice-frequency amplifiers and *4C* transistors for the carrier amplifiers. The first transistor is connected as a common collector and the second as a common emitter. By using them in this manner it is possible to employ the same type of transistor in both stages. Feedback is obtained by using high-side hybrid coils at both the input and output of the circuit in much the same manner as for vacuum-tube circuits. One significant difference is that in this transistor circuit only the second transistor introduces a 180-degree phase shift. This permits both input and output coils to return to a common ground as in a 3-tube vacuum-tube circuit and thereby avoid the circuit complications of a 2-tube circuit where one of the coils must float off ground.

The control signal for the variolosser is derived from the output of this amplifier and is in turn amplified by a 2-transistor feedback amplifier before being rectified. The characteristics of the compressor and associated expandor are shown in Fig. 3.

The modulator is a silicon diode ring with balance to eliminate the voice components. As this also effectively eliminates the carrier, it is reintroduced

---

Paper 56-242, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing January 12, 1956.

E. K. EBERHART, F. J. HALLENBECK, and E. H. PERKINS are with the Bell Telephone Laboratories, Inc., Lawrence, Mass.

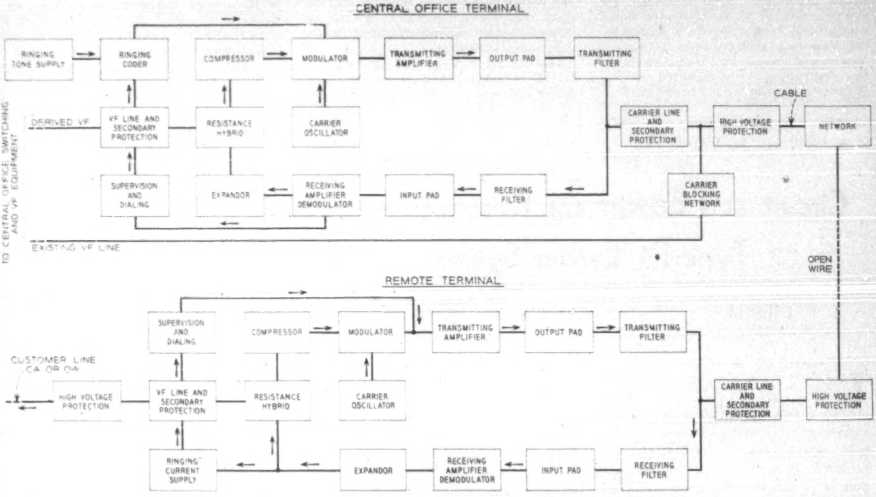

CENTRAL OFFICE TERMINAL
RINGING TONE SUPPLY
RINGING CODER
COMPRESSOR
MODULATOR
TRANSMITTING AMPLIFIER
OUTPUT PAD
TRANSMITTING FILTER
CABLE
DERIVED VF
VF LINE AND SECONDARY PROTECTION
RESISTANCE HYBRID
CARRIER OSCILLATOR
CARRIER LINE AND SECONDARY PROTECTION
HIGH VOLTAGE PROTECTION
NETWORK
SUPERVISION AND DIALING
EXPANDOR
RECEIVING AMPLIFIER DEMODULATOR
INPUT PAD
RECEIVING FILTER
CARRIER BLOCKING NETWORK
TO CENTRAL OFFICE SWITCHING AND VF EQUIPMENT
EXISTING VF LINE
OPEN WIRE
REMOTE TERMINAL
SUPERVISION AND DIALING
COMPRESSOR
MODULATOR
TRANSMITTING AMPLIFIER
OUTPUT PAD
TRANSMITTING FILTER
CUSTOMER LINE CA OR OW
HIGH VOLTAGE PROTECTION
VF LINE AND SECONDARY PROTECTION
RESISTANCE HYBRID
CARRIER OSCILLATOR
CARRIER LINE AND SECONDARY PROTECTION
HIGH VOLTAGE PROTECTION
RINGING CURRENT SUPPLY
EXPANDOR
RECEIVING AMPLIFIER DEMODULATOR
INPUT PAD
RECEIVING FILTER

Fig. 1. Over-all block diagram of P1 carrier telephone system

at the desired level for transmission at a hybrid connection following the modulator.

The carrier oscillator is a 2-transistor circuit. A *4C* transistor is used in a common-emitter circuit with a tank circuit in the collector circuit. Feedback is obtained from the tank circuit to a transformer to the base of the transistor. The output is obtained from the tank circuit through another common-emitter *4C* transistor as an amplifier.

The transmitting amplifier is used to raise the side-band level to the nominal value of −8 decibels (db) for transmission. It provides about 50 db of gain and has a characteristic as shown in Fig. 4.

The transmitting and receiving filters serve to connect the several channels of a system to the line. They pass only the frequencies desired for their particular channel, and prevent unwanted modulation products from the channel modulators from reaching the line and becoming crosstalk into other channels. And, vice versa, they exclude unwanted frequency signals from the demodulator where they would be demodulated as crosstalk. These filters and the carrier oscillators are the only frequency-conscious elements in the system. Typical loss characteristics for them are shown in Figs. 5 and 6.

The output pad permits the level of the transmitted signals to be adjusted for proper line co-ordination. The input pad permits the received signals to be adjusted for varying line losses.

The receiving amplifier is a 3-stage transistor regulating amplifier. It uses a *4C* transistor in a common-emitter circuit for the first stage followed by a modified *P*-terminal 2-stage amplifier. Feedback is necessary on the last two stages only to provide adequate control of the output impedance and the amplifier modulation. The input stage has sufficiently low signal level so that modulation is not a problem. The input impedance is also under sufficient control without resorting to the difficulties of a 3-stage feedback circuit. Because of the addition of the third transistor, the feedback is of the shunt type at the interstage between the first and second transistors.

The regulation in the receiving amplifier is of the backward-acting type. It uses silicon diodes at the amplifier input in a shunt-type variolosser. These diodes are activated by direct current from the control circuit. The control power is derived from the output of the receiving

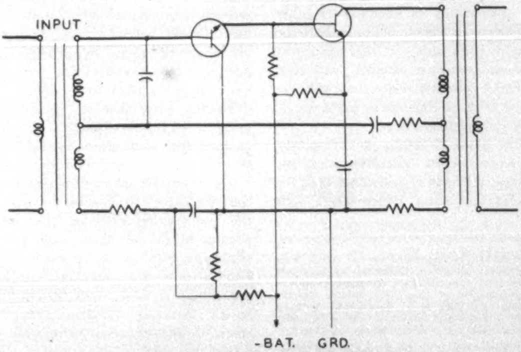

INPUT
OUTPUT
-BAT.
GRD.

Fig. 2. Transistor transmitting amplifier circuit

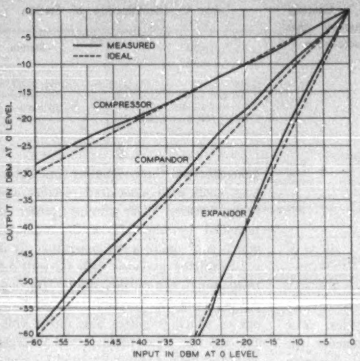


Fig. 3. (left). Compandor characteristics

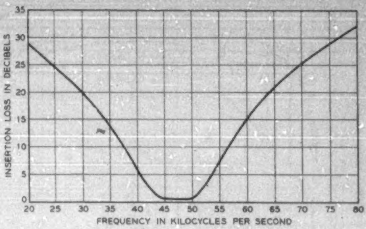


Fig. 5. (right). Transmitting filter - frequency characteristics

amplifier, amplified by a common-collector transistor amplifier and rectified by a voltage-doubler circuit. The reference to determine the regulator level is obtained from a reverse breakdown silicon diode of about 6 volts. The level of circuit regulation is controlled by potentiometer adjustment across the reference diode.

The demodulator is a bridge-type silicon diode rectifier. This rectifier is followed by a loss-pass filter to reject all but voice-frequency signals.

The expandor is of the forward acting type which derives control of the variolosser from the input signal. This signal is amplified, rectified by a voltage doubler, and bucked against a potentiometer-adjusted portion of the system battery. The variolosser uses silicon diodes in a series-type circuit. The variolosser is followed by a 2-transistor feedback amplifier similar to that used in the compressor to provide the necessary voice signal level.

The transmission qualities of this system are comparable to those obtained with toll-type carrier systems. The channel frequency characteristic is shown for each direction of tranmission in Fig. 7. Shown also for comparison are the limits prescribed for toll-system quality, which indicate this system meets toll requirements. The talking circuit has the good signal-to-noise ratio customarily found in carrier circuits. In service these channels will be found to be very quiet compared to the normal voice-frequency channels. The lower noise level is inherent in the use of carrier frequencies for transmission and is also aided more than 20 db by use of compandors.

Signaling

The signaling requirement for the system is that it shall fit into a voice-frequency circuit without needing any change of the existing signaling equipment at either end. This means that the central office terminal shall be activated by 20-cycle ringing which the remote terminal shall reproduce, and that the remote terminal shall be activated by dial pulsing, which the central office terminal shall reproduce. Thus, the two terminals have completely different signaling circuitry, and each direction of signaling is entirely separate. These terminals are shown in block schematic form in Fig. 8.

Central Office Signaling

The signaling by the central office is accomplished as in a voice circuit by putting a positive or negative d-c bias voltage superimposed with a 20-cycle a-c tone on either tip or ring to ground of the transmission pair. The different combinations of these variables provide the necessary control to select any one of four subscribers and is the information that must be reproduced by the remote terminal for transmission to the subscriber's line.

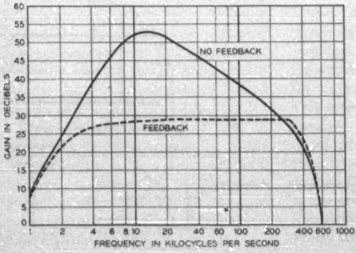


Fig. 4. (left). Transmitting amplifier characteristics

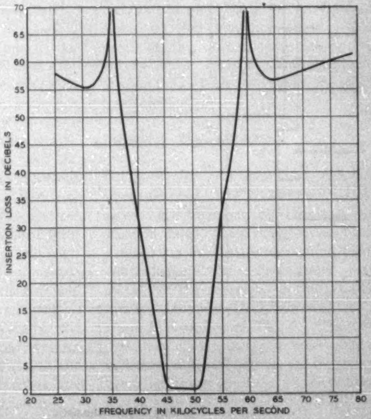


Fig. 6. (right). Receiving filter-frequency characteristics

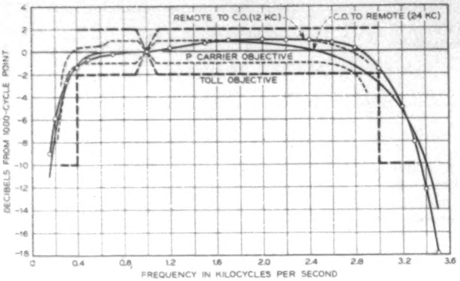


Fig. 7. Channel-frequency characteristics

Table I. Identification of Station

Tones	Station
2,500	ring party +
2,500 and 1,150	ring party −
2,500 and 1,750	tip party −
2,500, 1,750, and 1,150	tip party +

The central office terminal equipment consists of three voice-frequency oscillators and three associated keyers. These keyers connect the oscillators in different combinations to the circuit and so vary the voice-frequency tones transmitted over the carrier system. One of these keyers is activated whenever the plus d-c bias voltage is applied to the tip or the minus d-c bias voltage is applied to the ring of the voice-frequency input circuit to the terminal. This keyer is called the "either-or" circuit and it controls the oscillator generating 1,150 cycles. The second keyer is activated if either plus or minus d-c bias voltage is applied to the tip of the input circuit. This keyer is called the "if" circuit and it controls the oscillator generating 1,750 cycles. The third keyer is activated whenever the 20-cycle a-c tone is applied between either tip or ring and ground of the input circuit. This keyer is called the "when" circuit as it controls the oscillator generating 2,500 cycles at a 20-cycle rate. The information of any particular ring is transmitted to the remote terminal by combinations of these three tones. The information to identify which station is to be rung is shown in Table I. The output of these oscillators is amplified by a 1-stage common-emitter transistor amplifier and applied to the input of the carrier modulator as a voice signal.

The remote terminal equipment consists of a 3,000-cycle oscillator as a ringing generator and three signaling tone detectors, each with associated rectifier and relay circuit. The ringing power is provided by the 3,000-cycle oscillator which is a square-wave generator. It is a 2-transistor oscillator with a 2-watt *6A* transistor in the output stage. Its output is rectified by two full-wave rectifiers, one of which produces plus direct voltage and the other minus direct voltage. The 20-cycle a-c power required for ringing is obtained by alternately connecting the plus or minus voltage to the line at a 20-cycle rate.

The signaling-tone detectors are connected to the transmission circuit at the output of the channel expandor at all times when the subscribers are "on-hook." Under this condition the talking path is disconnected to increase the response of the detectors.

The signal tones as received at the remote terminal are amplified by three 1-stage common-emitter transistor amplifiers in parallel, with a tuned circuit in the output of each amplifier to select its

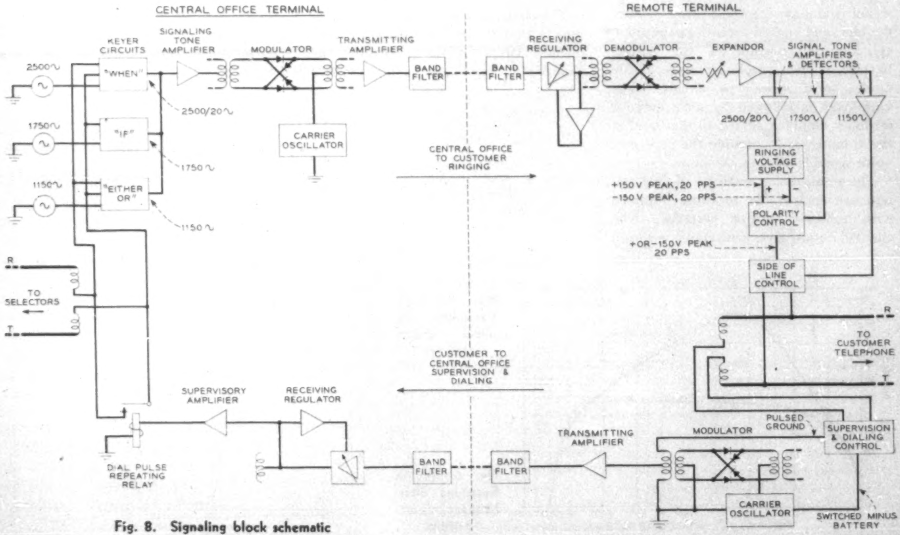

CENTRAL OFFICE TERMINAL
REMOTE TERMINAL
KEYER CIRCUITS
SIGNALING TONE AMPLIFIER
MODULATOR
TRANSMITTING AMPLIFIER
BAND FILTER
BAND FILTER
RECEIVING REGULATOR
DEMODULATOR
EXPANDOR
SIGNAL TONE AMPLIFIERS & DETECTORS
2500∿
"WHEN"
2500/20∿
1750∿
"IF"
1750∿
1150∿
"EITHER OR"
1150∿
CARRIER OSCILLATOR
CENTRAL OFFICE TO CUSTOMER RINGING
2500/20∿
1750∿
1150∿
RINGING VOLTAGE SUPPLY
+150 V PEAK, 20 PPS
−150 V PEAK, 20 PPS
+
−
POLARITY CONTROL
+OR−150V PEAK 20 PPS
SIDE OF LINE CONTROL
R
T
TO SELECTORS
R
T
TO CUSTOMER TELEPHONE
CUSTOMER TO CENTRAL OFFICE SUPERVISION & DIALING
SUPERVISORY AMPLIFIER
RECEIVING REGULATOR
DIAL PULSE REPEATING RELAY
BAND FILTER
BAND FILTER
TRANSMITTING AMPLIFIER
MODULATOR
PULSED GROUND
SUPERVISION & DIALING CONTROL
CARRIER OSCILLATOR
SWITCHED MINUS BATTERY

Fig. 8. Signaling block schematic

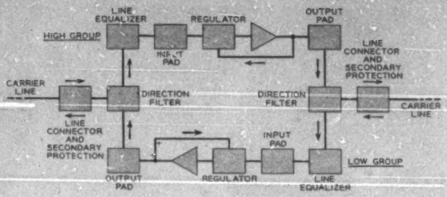


Fig. 9. Repeater block schematic

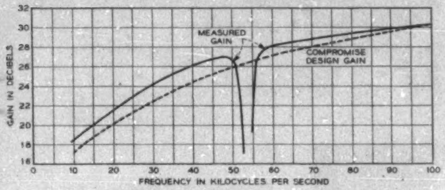


Fig. 10. Repeater gain-frequency characteristics

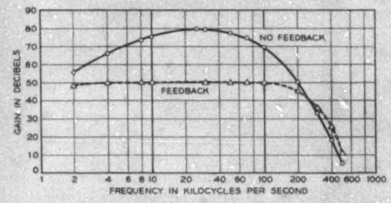


Fig. 11. Amplifier gain-frequency characteristics

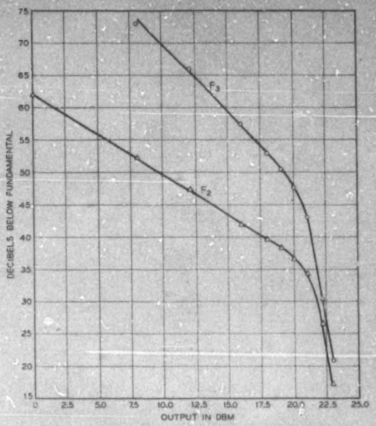


Fig. 12. Repeater modulation characteristics

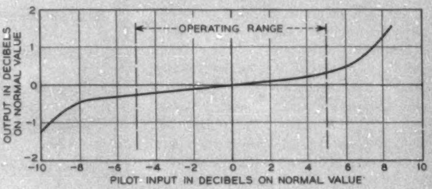


Fig. 13. Repeater regulation characteristic

proper tone. This tone is then rectified by a half-wave silicon-diode rectifier and the resulting direct current is applied to the associated mercury relay. The 1,150-cycle signal tone activates the relay and then connects the plus bias voltage to the tip or the minus bias voltage to the ring of the line. When this relay is not operated, minus bias voltage is applied to the tip or plus bias voltage is applied to the ring. The 1,750-cycle signal tone activates the relay that grounds the ring of the line. When this relay is not operated, the tip of the line is grounded. The 2,500-cycle signaling tone activates the relay that alternates at a 20-cycle rate the plus or minus ringing voltage that is applied between ground and the nongrounded side of the line.

Thus, the proper choice of plus or minus bias voltage, the creating of the 20-cycle ringing voltage, and the application of them between the tip or ring and ground is determined by the reception over the carrier circuit of the three in-band signaling tones.

Subscriber Signaling

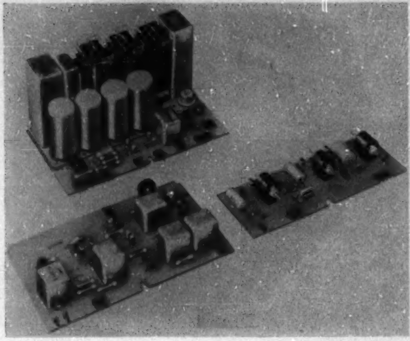
Two types of signaling are transmitted from the remote to the central office terminal. These are supervisory and dialing.

The subscriber in removing the handset establishes battery on the line or an "off-hook condition." This off-hook condition operates relays to activate the transmitting amplifier and to remove a short-circuit on the input of it. The remote terminal carrier is then transmitted to the central office. At the central office terminal the received carrier is amplified by the receiving amplifier, and a part, needed to activate the signaling equipment, is obtained from the output of the receiving amplifier. This output is amplified by a 2-transistor amplifier, then rectified, and the resulting direct current is applied to a relay which establishes a d-c path in the line to the central office. This operation recreates in the central office signaling circuit the line short circuit of the subscriber and brings in the supervisory signaling.

When the subscriber dials, he opens and short-circuits the subscriber line at each dial pulse. The opening of the line operates the relay that short-circuits the transmitting amplifier input but does not disconnect the amplifier power. The relays in the central office follow the pulsing of the received carrier, recreate the dial pulses there, and excite the selector equipment.

Repeater Transmission

The *P1* carrier system uses repeaters whenever the line transmission loss exceeds the maximum the terminals can accommodate. These repeaters use transistors for gain in place of vacuum tubes but otherwise are schematically very much like the repeaters of the type-*C5* carrier system. A block schematic for a repeater is shown in Fig. 9. The two directions of transmission have similar functions, differing only in the frequency band that is amplified and in the options that are used in the gain-regulating cir-



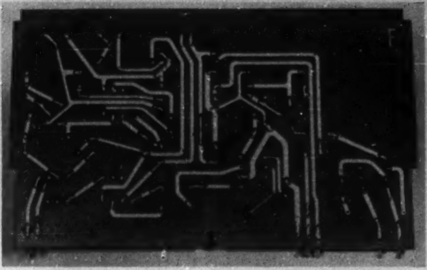


Fig. 14 (left). Typical coded networks using printed wiring board

Fig. 15 (above). Wiring side of typical printed wiring board

cuits. These options adapt it to the conditions of a particular system. Systems containing a single repeater may not require regulation, in some cases, depending on the plant layout.

The regulation in the outward direction from the central office terminal uses the integrated power of the channel carriers to determine the desired repeater gain. The regulation in the inward direction to the central office terminal use, a pilot frequency to control the regulation, as the channel carriers are present only when the channel is in use.

The repeater has identical high-low pass directional filters at each end of the repeater to separate the directions of transmission into the high- and low-frequency groups. When more than one repeater is used along a line, these directional filters have optional phase-correction sections to prevent side-band demodulation bucking. The directional filters are connected to the line through the same line-matching coils and secondary protection circuits used for the terminals. Between the drop sides of each pair of filters there is an input adjusting pad, an equalizer, a regulator, an amplifier, and an output adjusting pad. These output pads permit all levels to be adjusted to within 1 db of the desired transmission level. The input pads are used to build out the line loss for a non-regulated repeater to its maximum value and for a regulated repeater to build out the line loss so as to center the repeater regulator. These pads are in 2-db steps up to 30 db.

The equalizer is designed to a compromise characteristic of the many types of transmission line that are usual in the rural plant. The design covers both groups with one unit so that identical equalizers are used for both the high-group and low-group sides of the repeater.

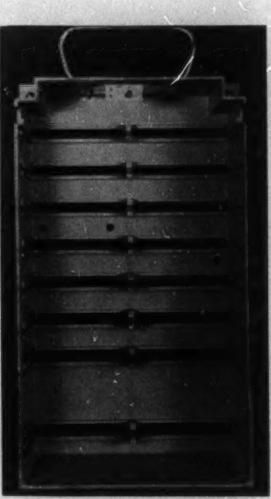


Fig. 16. Prototype model of connector box unequipped showing grid wires

The repeater regulators are backward-acting devices using a shunt-type variolosser in front of the repeater amplifier; they are controlled by the signal at the amplifier output. The regulator control circuit in the outward direction from the central office terminal is not frequency-sensitive and integrates the power of whatever channel carriers are present. This power is then amplified by a transistor amplifier, rectified by a voltage doubler, and applied to the variolosser to provide the loss. The rectifier has an average voltage type of response so that the regulator loss will not respond to the rapid fluctuations of power in the channel side-bands which are 9 db below the carriers. The reference to determine the regulator level is obtained, as for the terminal regulators, from a reverse breakdown silicon diode of about 5 volts.

The regulator control circuit in the inward direction to the central office terminal is frequency-sensitive. It uses a crystal filter at the input of its control circuit so that only the pilot frequency will activate the control circuit. This regulation uses a peak signal rectifier as only the pilot frequency is present. The regulator control amplifiers use transistors in common-emitter circuits, with sufficient feedback to stabilize their gain.

The repeater amplifiers are designed to cover both groups so each repeater contains two identical amplifiers. The amplifier is a 3-transistor amplifier with each stage connected as a common emitter. It uses p-n-p type-*7B* and *6B* transistors in the first and last stages respectively, and an n-p-n type-*4C* transistor in the second stage. Local feedback is required around each transistor to reduce the gain spread and phase variations among units. Over-all feedback is obtained around the three transistors with high-side hybrid coils at input and output. Output pads similar to the input pads are used at the repeater output to permit co-ordination of signal levels among joining systems on a multiple circuit line.

The characteristics of a typical repeater are shown in Figs. 10 through 13. These characteristics are taken from preliminary laboratory design data but are indicative of the final product. The repeater gain frequency characteristic, Fig. 10, is shown plotted with the design objective. There is significant departure only at the cut-apart frequencies. This will be

Vol.
Pt. 1
1956

Fig. 17 (above) Close-up of terminal connectors

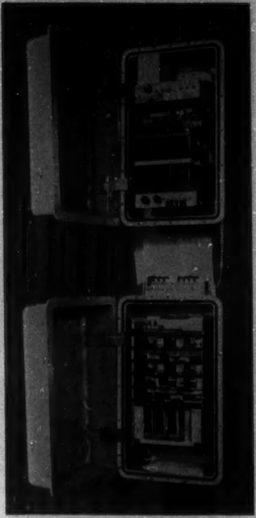


Fig. 19. Pole-mounted terminal and power supply with apparatus cases open

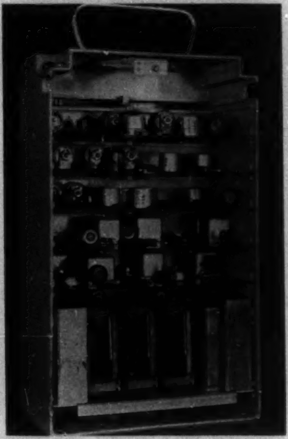


Fig. 18. Typical terminal networks mounted in a prototype connector box

corrected sufficiently to permit as many as five repeaters to be used in tandem. As this objective is a compromise of the loss among several types of lines that are standard for this system, the departures in system use will vary considerably from these shown.

The repeater amplifier gain frequency characteristic, Fig. 11, shows the nonregenerative peak gain of the 3-stage transistor amplifier of 80 db, and the feedback gain characteristic of 50 db. This provides 20 db or more of feedback over the transmitted band to produce the necessary operating stability with temperature and battery variations and a working value of modulation suppression.

The repeater modulation characteristic, Fig. 12, indicates that the repeater will operate in the system with overload margins very similar to those in the type-*N* system which has proved satisfactory.

The regulation characteristic, Fig. 13, indicates a 15-db range with an input-output stiffness ratio of about 10. This provides an output constant within 1/2 db of line-up value. This characteristic is a design objective that will be closely met. To date, sufficient data are not available to plot a design performance curve.

Power Requirements

One of the unique requirements for this system was that the power consumption be an absolute minimum especially for the remote terminal. The use of transistors has aided greatly in that endeavor. Their use has also eliminated the need for the 130-volt battery required for vacuum tubes. As transistors do not require any warm-up time, the circuitry has been arranged so that, in the remote terminal, the compressor, carrier oscillator, transmitting amplifier, and supervisory relay consume power only when the subscriber is off-hook. These arrangements produce a remote terminal requiring only 0.6 watt of power during the idle period and only slightly more than 2 watts during the off-hook talking period. The majority of this off-hook drain is the transmitter current required by the subset. The ringing period requires about 8 watts additional power but is on less than 1/2 per cent of the time and so represents very little long-time drain.

The similarity of transmission circuitry in the two terminals has also produced a central office terminal with a very low power drain. It is about 2 watts. It is more than the remote terminal because the apparatus at the central office is not turned off during idle periods as the added circuitry is not warranted where ample power is available.

These low power consumptions indicate a real economical system. The measure of how well the low power design objective has been realized is shown in Table II which lists the powers required by several parts of this system and those

Table II. Power Requirements, Watts

Unit	Type P1	Type N
Compressor	0.28	3.5
Transmitting amplifier	0.13	8.0
Carrier oscillator	0.20	6.0
Receiving amplifier	0.21	13.0
Expandor	0.20	13.0
Repeater amplifier with regulator	3.30	23.0

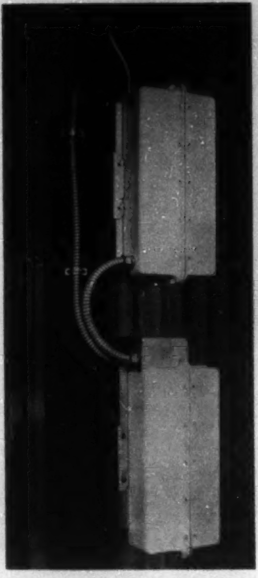


Fig. 20. Pole-mounted terminal and power showing connecting wiring

Fig. 21. Test set

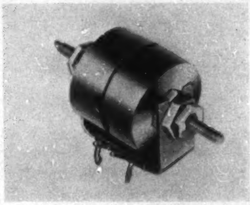


Fig. 22. Ferrite inductor

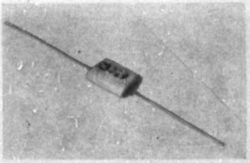


Fig. 23. Mylar capacitor

of the type-*N* system which function alike, except that one uses transistors and the other vacuum tubes. Further description of the power arrangements is given in reference 1.

Equipment

As mentioned earlier under the broad objectives of the system, the equipment for a rural carrier system must have low cost and provide ease of maintenance, particularly at remote locations. This has been accomplished for the *P1* carrier with plug-in units using printed wiring with miniaturization of components as necessary.

The first step in the equipment design was to break down the terminal and repeater circuits into logical sections and to assign a network code to each section. The terminal for example has the following breakdown of networks:

1. Line connector.
2. Oscillator-modulator and transmitting filter.
3. Receiving filter (half board).
4. Transmitting amplifier (half board).
5. Receiver amplifier and demodulator.
6. Compressor.
7. Expandor.
8. Signaling-tone amplifier and detector (remote).
9. Signaling-tone oscillator (central office).
10. Signaling and voice-frequency circuit.

With the view towards mechanized assembly, printed wiring circuitry has been employed. This also simplifies assembly, testing, and inspection and assures a more uniform product. With the exception of a few heavy components such as relays and large transformers, all apparatus components are mounted directly on printed wiring boards and held in place by their own leads inserted into holes in the printed wiring boards and soldered directly to the printed wiring "lands." A land is the area on the leads where connections are made.

Each coded network consists of its apparatus components mounted on a printed wiring board. The boards are equipped with a unique design of terminal connector, described and illustrated in this article, which permits them to be easily plugged into a wire grid assembly.

The transmitter amplifier and receiver-amplifier demodulator are frequency-conscious and, therefore, have a code for each assigned frequency, i.e., 12, 24, 36, 48, 60, 72, 84, and 96 kc. Thus a plug-in board of the desired frequency can be readily inserted or replaced as required by job conditions.

Fig. 14 shows a group of typical networks as viewed from the apparatus side. The major boards measure 8 by 5 inches and are made of 1/16 phenol fibre with 2-ounce laminated copper on one side only. For small networks half boards are used, two being clamped together before insertion into the connector. Fig. 15 shows the printed wiring side of a typical network with the code number, name, and mounting position letter (*F*) etched with the circuitry, and tinned prior to soldering the components in place. With the view towards automation, the pigtail components and straps on the boards are laid out in definite patterns on predetermined modules.

The networks which go to make up a terminal or a repeater for one channel are mounted in an inner box known as a connector which will accommodate nine networks. The connector is a cast phenolic box with grooves for retaining the

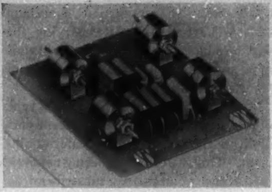


Fig. 24. Receiving channel filter

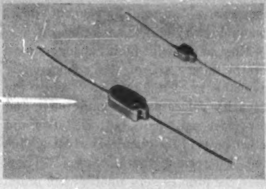


Fig. 25. Tantalum capacitor

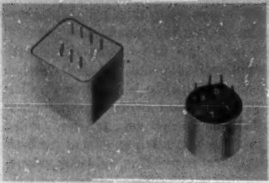


Fig. 26. Carrier-frequency and voice-frequency transformers

boards. The groove positions are identified by letters starting with *A* at the top of the connector. Fig. 16 shows a prototype model of the connector. The rear of the connector consists of a grid of 36 accurately spaced 22-gauge copperweld wires which are the means of making connections between networks. A connector box measures 15 by $8^{1}/_{2}$ by $5^{3}/_{4}$ inches. Connections to the grid are accomplished by the use of terminal connectors fastened to the boards. Each terminal connector consists of a pair of bifurcated cantilever springs made of phosphor bronze. Fig. 17 shows a close-up view of the terminal connectors. These terminals are used at the front of the networks as test connectors and provide a means for hooking on a portable test set. A protective coating on the grid wires and terminals affords better wearing quality.

The chief advantage of the foregoing design is the ease with which an entire terminal or repeater can be carried to an installation or a maintenance depot where it can be quickly installed or replaced in its entirety, or by networks, as required. Also, as brought out earlier, the frequency assignments for a line can be quickly changed by plugging in networks of desired frequencies into the terminals, and the expanded or changed traffic conditions are ready to be served.

A complete terminal, except for the line network, is illustrated by Fig. 18. The bottom position in this terminal contains the network for the signaling circuit which employs wired-in conventional relays as well as printed wiring circuitry. The top position, shown vacant in this figure, is where the line network mounts. The line network connects the terminal to the line and power supply through a local cable connected to screw terminals on the network. The outer end connects to a gasketed terminal block at the top of an outer housing, known as an apparatus case, at remote locations. Carbon block protectors are built into the terminal block for the line leads. A secondary protection circuit, referred to earlier, consisting of a series resistor and a pair of diodes connected back-to-back in shunt across the line is provided at remote locations. The latter is located on the line network. Also on these networks are mounted input and output pads which can vary the levels. These pads are fastened to screw terminals and are readily interchangeable. They are available in 2-db steps up to 30 db.

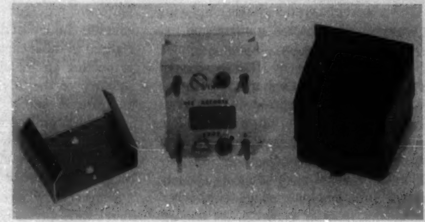


Fig. 27. Carrier blocking network

Fig. 19 shows a pole-mounted terminal with its associated power supply mounted above it, both being mounted in apparatus cases of the same size. These cases have an over-all measurement of $12^{1}/_{2}$ by $23^{3}/_{4}$ by $10^{5}/_{8}$ inches.

The apparatus case is made of die-cast aluminum and is built of two similar halves. A round O-ring neoprene gasket is used to seal the two halves to keep out moisture. A 10-ounce bag of silica-gel is placed in the bottom of each case to keep the moisture content of the trapped air to a minimum. The two halves are secured together by 16 screws to assure even pressure on the gasket. The case is painted white to reflect the heat from the sun. Sun shields may be used where extreme temperatures prevail as a further means of keeping the temperature inside the case to a minimum.

The terminals in central office, are arranged to mount in pairs on a 19 inch relay rack and, with the associated terminal strips and carrier blocking networks (when required), they occupy 21 inches of vertical mounting space. They may also be wall-mounted.

The line network for both central office and remote locations can be removed from the connector box and stored in a shelf located directly under the connector without disconnecting the local cable. This permits the connector and its remaining networks to be removed as a unit. The shelf position can be seen in Fig. 19.

Repeaters mount the same as terminals in remote or central office locations. At a remote location the terminal or repeater is mounted on the pole to be accessible from the ground for convenient maintenance. Conventional saddles are used to mount the apparatus cases to the pole, as shown in Fig. 20. This figure also illustrates the use of flexible conduit known as "greenfield" or "flexsteel" to protect physically the external wiring.

With the arrangement described there is no need for any soldering operations at remote locations since the plug-in units can be readily replaced because of failure of components, or an entire terminal or repeater can be carried in its connector case and be substituted as a complete unit by simply storing the line network and loosening two screws at the top of the connector case. Testing of the equipment is accomplished in the field by the use of a specially designed unit, as shown in Fig. 21.

Components

Development of the passive components of the *P1* carrier system including the various filters and other networks were influenced by three major considerations. Again, the manufacturing cost had to be as low as possible, though the traditional standards of Bell System service life had to be maintained. The components had to lend themselves to maximum utilization of the printed wiring techniques to be used as the basic equipment method. And lastly, advantage wherever possible was to be taken of the fact that transistors are low-power devices.

Componentwise, filters are the most important single assembly determining the first cost of a carrier system employing frequency division multiplexing and frequency separation for obtaining equivalent 4-wire operation on the line. Experience has shown that one-quarter or more of the first cost is chargeable to the various filters. Much attention was therefore paid to this problem even in the system planning stage. The decision to employ double-side-band modulation was largely based on the knowledge that, when frequency space is available, the double-side-band channel filters are generally less expensive.

With this decision made, every effort was directed toward the achievement of channel filter designs of maximum efficiency in element utilization. Advantage was taken of the more leisurely rising characteristics of double-side-band filters. Inexpensive wide-limit capacitors are employed, and the desired performance is achieved through the use of an adjustable ferrite inductor expressly developed for the *P1* carrier system. The filters are readily adjusted in the manufacturing process with the use of visual display testing circuits.

The inductor which makes this possible is shown in Fig. 22. It is designed for printed wiring use and provides a wide range of inductance change while maintaining excellent Q performance in the carrier and voice range. This is accomplished in a single basic design by so selecting the winding for particular nominal inductances that the air-gap adjustment remains at or near its most efficient setting. Inductors of this type are used not only in all the channel, demodulator, and signaling filters in the terminal and directional filters at the repeater but also in the channel oscillators and other parts of the circuitry where an inexpensive, adjustable element offers manufacturing or field advantages.

Most of the wide-limit capacitors used in the filters are of the commercially available molded-mica type. Where the capacity values would require large and expensive mica units both in filters and other parts of the circuit, newly available foil-Mylar capacitors are used. These take the form of very small pigtail units, as shown in Fig. 23.

The capacitors have low working voltages in these miniature sizes and can be employed because of the low-voltage protection provided for the transistorized circuits. Both cost and space savings are realized in these capacitors since no cans or potting are required because of the stability of the Mylar dielectric under moisture exposure.

A typical printed wiring assembly of a receiving channel filter is shown in Fig. 24. This unit employs all three of the basic components described in the foregoing.

Another new type of capacitor has found wide-spread use in the *P1* carrier system. This is a solid tantalum electrolytic capacitor used in place of the usual paste or liquid electrolytic capacitor. The solid electrolyte is manganese dioxide deposited upon the capacitor surfaces. The anode is made from tantalum metal and upon its surfaces is deposited the tantalum oxide which forms the dielectric. The cathode is an enveloping metal completing the capacitor structure. The new design of capacitor provides up to 100 microfarads in a very small volume. It is expected to be less expensive than other electrolytic capacitors while at the same time providing a rugged structure, fairly inert electrochemically, with better stability in operation and storage. Fig. 25 shows typical solid tantalum units. Transformer needs in the *P1* carrier system are provided for with two miniature structures which were made possible by the use of low-power transistor circuits.

The carrier-frequency units employ a manganese zinc ferrite core, a spool winding, and wire terminals which permit assembly on printed wiring boards. They are potted with an asphalt compound in a cylindrical aluminum can.

The voice-frequency transformers are wound on laminated core structures of either *45* permalloy or molybdenum permalloy. The units are potted in an epoxy resin in rectangular aluminum cans. The terminal plate carrying the wire terminals for mounting is a cast unit of a styrene polyester. Both types of transformer are shown in Fig. 26.

Also deserving mention is a special type of line network which does not form part of either the terminal or repeater equipment. This is a carrier-blocking network used at each point where a subscriber drop is made from a carrier line. By presenting a high impedance to carrier frequencies, the network prevents absorption of carrier power in the drops. It also prevents any harmonics produced at the subset from interfering in the carrier circuit. Several electrical varieties are available to provide for the various lengths of subscriber drop.

The physical arrangement is shown in Fig. 27. The networks are cast in a styrene polyester. High-voltage protection is self-contained and sturdy terminals are provided for bridle-wire connection. By means of side slots in the casting, the network may be mounted on a wedge-shaped holder which is fastened to the pole. A flexible rubber cover is snapped over the face of the network to protect against deleterious weather effects.

Conclusion

The design objectives for this system required a new approach to almost every phase of the development. The use of transistors introduced sweeping changes in the detailed circuitry and created the demand for many new component parts. The use of printed wiring introduced a new necessity for close co-operation between design and manufacturing objectives, and the desire for low cost was an important factor. The necessity for field flexibility and ease of maintenance required extensive broadening of the training and handling concepts. All of these factors have been blended into this design, and its introduction into the telephone plant will give another indication of the use of carrier systems to provide good quality of communication service at low cost.

Reference

1. Power Supplies for the Type-P1 Carrier System, D. H. Smith. *AIEE Transactions*, vol. 75, pt. I, May 1956, pp. 191–95.

Application of the Type-P1 Carrier System to Rural Telephone Lines

J. D. HOWARD, JR.
MEMBER AIEE

THE rural service situation is one of the most pressing and difficult problems to solve in the telephone industry. The use of fixed outside plant such as cable, open wire, or *B* rural wire (6-pair distribution wire or cable) proves expensive in many instances for providing each customer in the rural area with the grade of service he desires when he wants it. This is particularly true for rural areas because of the difficulty of accurately forecasting rural service demands.

The advances in the electronic field in the past decade have resulted in improved carrier equipment at substantially reduced cost so that shorter and shorter telephone circuits have been provided economically by carrier. The use of carrier systems in the trunk and toll fields is already paying off in economy and flexibility by providing circuits with movable equipment and a minimum of fixed outside plant. Therefore, if sufficiently low-cost carrier equipment can be obtained and used as a general method for supplying rural circuits, it should provide similar economy and flexibility since the capacity of any line or cable could be expanded and contracted at will as the demands occur. This concept has been the basis for the development of rural carrier systems.

The transistor, with its associated low-power consumption which allows the use of low-voltage and subminiature size capacitors, resistors, and inductors that can be interconnected on printed wiring boards and are adaptable to automatic factory assembly, point toward carrier equipment of sufficiently low cost to be attractive for providing rural telephone circuits. Therefore, transistors and printed wiring assemblies have been incorporated in the *P1* carrier system design.

The first requirement for the *P1* carrier system as for any rural carrier system is low cost but, unless the system can be applied to existing base plant by means of simplified engineering of carrier line layouts and installation and can be easily maintained, and other papers[1-2] show how the *P1* carrier system has been designed to enable a simplified approach to carrier line layout, installation, and maintenance to meet the over-all requirements for an economical rural subscriber carrier system.

Application of P1 Carrier

The *P1* carrier system is to be applied to normal exchange base plant facilities engineered in accordance with the resistance design methods as covered in a paper by Bogan and Young.[3] These facilities consist of mixed gauges of high-capacitance cable extended at their outer extremities by *109* steel, *104* copper, or copper-steel wire. A typical rural plant and the problems involved in carrier application on this type of facilities were described in a paper by Hochgraf and Watling.[4] A review of both of these papers will be of considerable value in understanding some of the basic concepts discussed in the following.

To engineer the carrier line design or carrier layout, the engineer must be familiar with the basic capabilities of the carrier system, but he need not have a complete knowledge of carrier-system operation. He is primarily concerned with the basic "building blocks" and the limitations that must be considered in applying the system to the physical line. The basic building blocks are as follows:

1. Central office terminals.
2. Remote terminals.
3. Repeaters.
4. A-c or d-c remote terminal and repeater power supplies.
5. Low-pass filters.
6. High-pass filters.
7. Carrier line termination.

Fig. 1 shows in block schematic form a carrier application of all of the various components listed in the foregoing. In addition to these, an autotransformer or junction line filter may be required at the junction of the open wire and cable. The autotransformer is required to eliminate reflection losses and to reduce crosstalk at carrier frequencies caused by impedance mismatch of the cable and open wire. The junction line filter is required to allow the carrier and the physical voice-frequency (VF) circuit to be on different pairs in the cable section, where the physical circuit is so long that load coils are required for VF and nonloaded cable pairs are required for the carrier, and the same open-wire circuit is used for both voice and carrier beyond the cable open-wire junction.

Fig. 1 gives an example of two types of subscriber distribution beyond the remote terminals referred to as wire distribution and filter distribution. Wire distribution is indicated by the VF extensions of channels 1 and 4, and filter distribution is shown for channels 2 and 3 and for the physical VF circuit.

The application of the *P1* carrier consists of determining the carrier line layout necessary to meet the over-all requirements for a suitable carrier transmission path on the available physical facilities. To do this adequately, the engineer must be given certain ground rules based on the characteristics of the carrier system.

The *P1* carrier system operates in the frequency range of 8 to 100 kc with three frequency arrangements, as shown in Fig. 2. The stackable frequency arrangement utilizes adjacent frequencies to provide for opposite directions of transmission and is used for nonrepeatered operation. The design of the carrier system is such that the bare line loss at the top frequency of each individual channel can be 30 decibels (db) from the central office terminal to the remote terminal, as shown in Fig. 3(A). The VF extension beyond the remote terminal can be 390 ohms (5 miles of *109* steel wire). The 390 ohms are controlled by the talking battery supply requirements of the subsets with the battery being supplied from the remote terminal power supply. The *P1* carrier system has been designed to operate with the *500*-type telephone set as this set insures better over-all performance because of its improved dialing, ringing, and transmission aspects. The over-all transmission of a carrier channel and its VF extension with the use of the *500*-type telephone set will be as good or better than that obtained on long rural lines with physical plant laid out by the afore-mentioned resistance design method.[3]

Paper 56-241, recommended by the AIEE Committee on Wire Communications Systems and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30-February 3, 1956. Manuscript submitted October 20, 1955; made available for printing December 13, 1955.

J. D. HOWARD, JR., is with American Telephone and Telegraph Company, New York, N. Y.

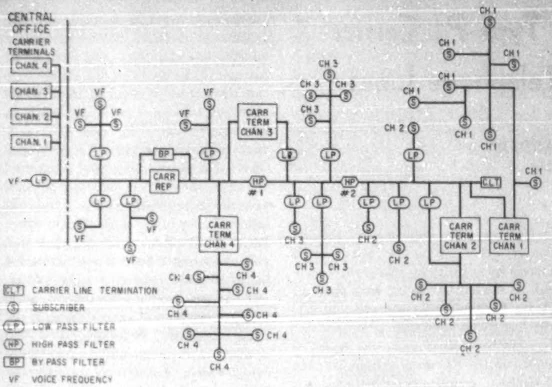

CENTRAL OFFICE CARRIER TERMINALS
CHAN. 4
CHAN. 3
CHAN. 2
CHAN. 1
VF
LP
BP
CARR REP
HP #1
HP #2
CARR TERM CHAN 3
CARR TERM CHAN 4
CARR TERM CHAN 2
CARR TERM CHAN 1
CLT
CH 1
CH 2
CH 3
CH 4
CLT CARRIER LINE TERMINATION
S SUBSCRIBER
LP LOW PASS FILTER
HP HIGH PASS FILTER
BP BY PASS FILTER
VF VOICE FREQUENCY

Fig. 1. Typical system application

The normal and staggered grouped frequency arrangement are for repeatered operation and the system design allows 30-db bare line attenuation at the top frequency (96 kc) either between the central office terminal and the first repeater, or between repeaters and between the last repeater and the remote terminal 30 db at the particular frequency involved with the individual channel. The repeatered system can use a maximum of four repeaters for a total line loss of 150 db at 96 kc. The repeatered arrangement is shown in Fig. 3(B). Referring to this figure, it can be seen that, of the four repeaters, a maximum of two are permissible in open wire with the last cable repeater at least 1 mile back in the cable from the cable open-wire junction. These requirements are based on noise and crosstalk objectives.

In addition to bare line loss, there will be approximately 3 db of miscellaneous losses in any normal channel layout including carrier blocking networks, high-pass filter losses, bridging losses of other terminals on the carrier line, cable and open-wire junction autotransformer losses, or cable and open-wire junction line filter losses. Since these losses do not add in phase, it is simpler to use an average loss factor to cover most conditions rather than to make computations to determine a definite loss for each set of conditions that might exist. Thus a total of 33 db including the bare line loss plus miscellaneous loss from terminal to terminal may be expected. The maximum system transmitting line level of each sideband is −8 db (referred to the channel 2-wire input 0 db level point). Carrier is transmitted 12 db above each sideband or +4 decibels referred to 1 millivolt (dbm) at the terminal output to the carrier line (i.e., −8-db system level point). The maximum transmitted carrier energy is +4 dbm at the central office line terminals and +0.5 dbm at the remote terminals on the carrier line, i.e., where the remote terminals are bridged to the line after experiencing about $3^1/_2$-db bridging loss. Therefore, the minimum carrier power in the central office to remote direction will be −29 dbm (+4 dbm −33 dbm = −29 dbm) at the bridging point of the remote terminal. In the remote-to-central-office direction, the minimum power will be −32.5 dbm (0.5 dbm −33 dbm = −32.5 dbm) at the line connection of the central office terminal.

The *P1* terminals provide 15 db of automatic channel regulation and the repeaters provide 10 db of automatic group regulation to counteract variations and loss caused by weather changes. Input and output pads at the terminals and input pads at repeaters are provided where necessary to eliminate the possibility of overloading the equipment by building the line section out to required loss.

In laying out the carrier line design, it is necessary to determine the possible locations for the remote terminals based on accessibility to the subscribers to be served and the availability of commercial power since this type of battery supply will be the most economical. Of course, when a-c power is not available, the primary air-cell batteries can be used. Having determined the ideal location of the terminals from a physical standpoint, the makeup of the physical circuits back to the central office must be determined and computations of the facilities attenuation at carrier frequencies have to be made to determine the location and number of repeaters required, if any, and to check for the necessity of input and output pads at the terminals.

The many variations in the exchange plant facilities such as a mixture of 26-, 24-, 22-, and 19-gauge cable would make a factor-computing method very laborious. This pointed to the need of a simplified method of determining the carrier line losses. This led to the development of length-loss charts so that a carrier line design could be made very similar to the loop cable design with the use of the resistance design methods as

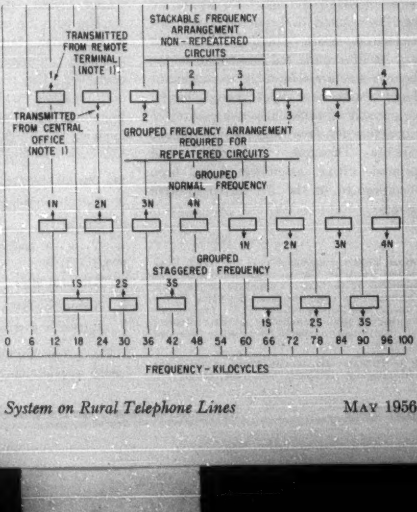
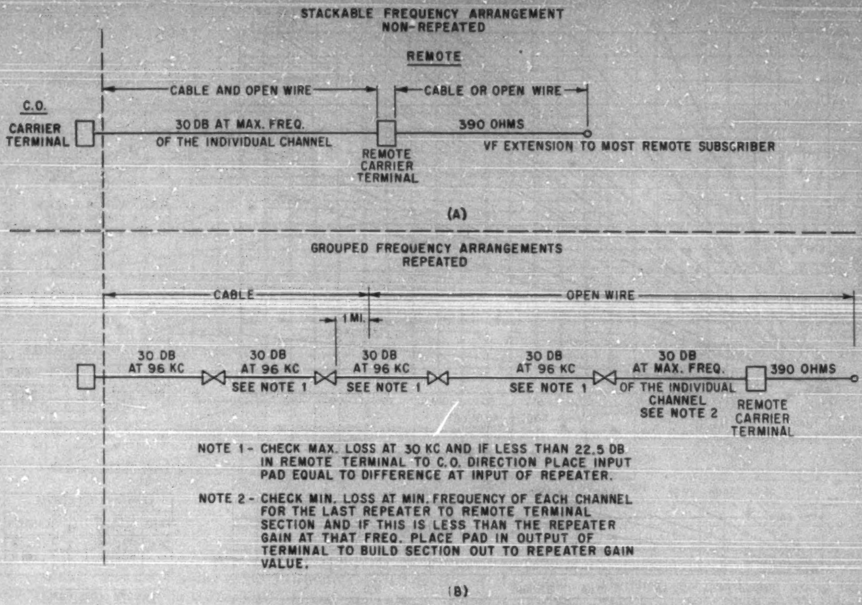

TRANSMITTED FROM REMOTE TERMINAL (NOTE 1)
STACKABLE FREQUENCY ARRANGEMENT NON-REPEATERED CIRCUITS
TRANSMITTED FROM CENTRAL OFFICE (NOTE 1)
1
2
3
4
GROUPED FREQUENCY ARRANGEMENT REQUIRED FOR REPEATERED CIRCUITS
GROUPED NORMAL FREQUENCY
1N
2N
3N
4N
GROUPED STAGGERED FREQUENCY
1S
2S
3S
0
6
12
18
24
30
36
42
48
54
60
66
72
78
84
90
96
100
FREQUENCY - KILOCYCLES

Fig. 2. P1-carrier frequency allocation

Notes:

1. Arrows indicate direction of transmission which should be used unless co-ordination is involved
2. Numeral indicates channel number
3. Letter along with number designates normal or staggered grouped frequencies
4. Sideband width, 3,200 cycles per second


STACKABLE FREQUENCY ARRANGEMENT
NON-REPEATED
REMOTE
C.O.
CARRIER TERMINAL
CABLE AND OPEN WIRE
CABLE OR OPEN WIRE
30 DB AT MAX. FREQ. OF THE INDIVIDUAL CHANNEL
390 OHMS
VF EXTENSION TO MOST REMOTE SUBSCRIBER
REMOTE CARRIER TERMINAL
(A)
GROUPED FREQUENCY ARRANGEMENTS
REPEATED
CABLE
OPEN WIRE
1 MI.
30 DB AT 96 KC
30 DB AT 96 KC SEE NOTE 1
30 DB AT 96 KC SEE NOTE 1
30 DB AT 96 KC SEE NOTE 1
30 DB AT MAX. FREQ. OF THE INDIVIDUAL CHANNEL SEE NOTE 2
390 OHMS
REMOTE CARRIER TERMINAL
(B)

NOTE 1 - CHECK MAX. LOSS AT 30 KC AND IF LESS THAN 22.5 DB IN REMOTE TERMINAL TO C.O. DIRECTION PLACE INPUT PAD EQUAL TO DIFFERENCE AT INPUT OF REPEATER.

NOTE 2 - CHECK MIN. LOSS AT MIN. FREQUENCY OF EACH CHANNEL FOR THE LAST REPEATER TO REMOTE TERMINAL SECTION AND IF THIS IS LESS THAN THE REPEATER GAIN AT THAT FREQ. PLACE PAD IN OUTPUT OF TERMINAL TO BUILD SECTION OUT TO REPEATER GAIN VALUE.

Fig. 3. Maximum bare line attenuation, P1 carrier system

mentioned in the foregoing.[3] Fig. 4 shows the 96-kc length-loss chart based on maximum facility loss (+140 degrees Fahrenheit loss factors for cable and wet or ice conditions for wire) which is used to lay out repeater spacings and channel-4 over-all circuit design. Figs. 5 and 6 are examples of the maximum and minimum loss charts for other carrier frequencies utilized. These charts are used for terminal-to-terminal section layouts of all but channel 4 and repeatered systems using the normal or staggered frequency arrangements.

In addition to information on repeater spacings and terminal spacings based on facility loss at carrier frequency, the engineer must know how many channels of carrier can be applied to each circuit of open wire, cable, or *B* rural wire. In regard to cable and *B* rural wire, crosstalk tests have indicated that four channels of either the stackable or normal grouped frequency arrangement can be used on each pair, or the three channels of the staggered grouped frequency arrangement can be used on each pair. However, in the case of open wire, the frequency arrangement and number of channels are very dependent upon the type of transposition system used. The most commonly used transposition system in rural lines of the Bell System is designated the *R1*. Table I gives the carrier assignments for the various frequency arrangements on a 1- and 2-crossarm route utilizing the *R1* transposition method.

Table I. Number of Carrier Channels that Can Be Assigned to the Open-Wire Circuits in Accordance with the P1-Carrier Frequency Arrangements

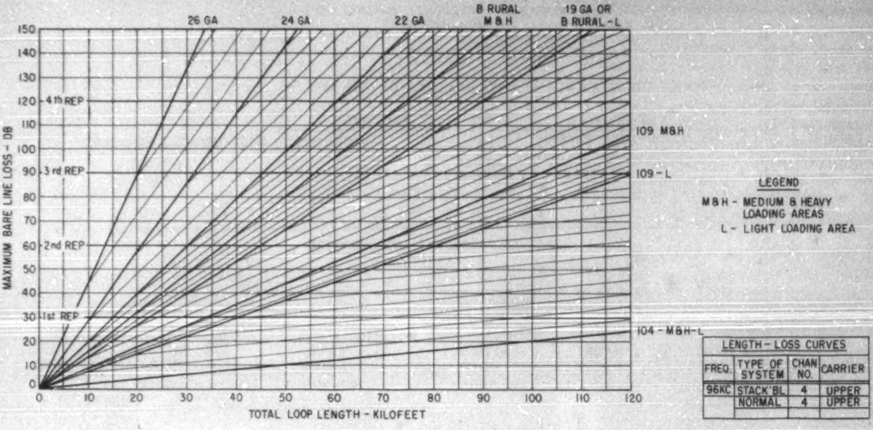
Pins	Stackable	Grouped
1 and 2	4	4*N**
3 and 4	1	3*S*†
5 and 6	0	0
7 and 8	4	4*N*
9 and 10	2	3*S*
Total for one crossarm	11	14
11 and 12	1	0
13 and 14	2	0
15 and 16	0	0
17 and 18	4	4*N*
19 and 20	4	0
Total for two crossarms	22	18

* *N*—normal frequency arrangement.

† *S*—staggered frequency arrangement; see Fig. 2.

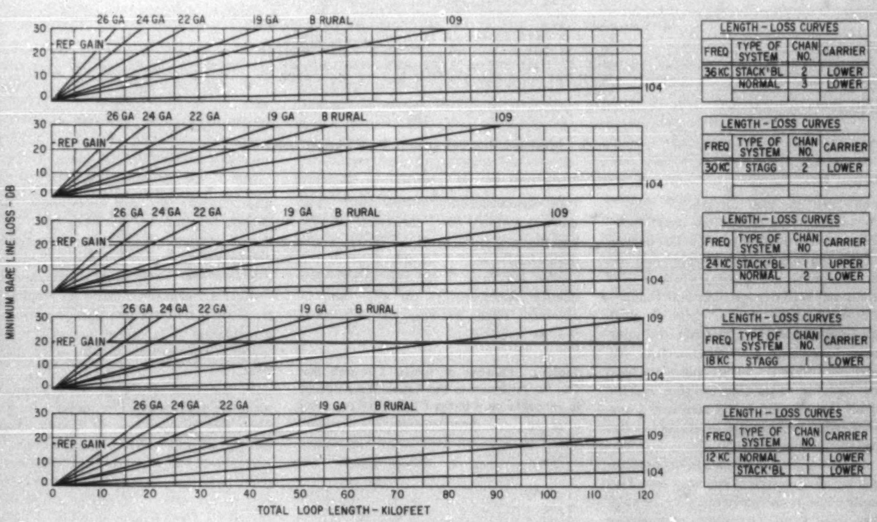
The stackable and grouped frequency arrangements do not co-ordinate and cannot be used on the same line. The group normal and group staggered arrangements do co-ordinate and can be used on the same line. Referring to Table I a 1-arm route using the stackable frequency arrangement will provide for a maximum of 11 carrier channels, and a 2-arm route can have a total of 22 channels.

In the case of the group systems, 8 channels of the normal frequency arrangement can be used on one arm or 12 on two arms and, with a combination of normal and staggered, a total of 14 channels can be used on one arm and 18 channels on two arms. It will be noted that the group arrangement provides 4 channels less on a 2-crossarm basis than the stackable arrangement which is primarily an economic consideration in that the grouped arrangement is used for repeatered operation with repeaters handling a complete system of three or four channels; this makes repeaters rather expensive for a single- or double-channel circuit assignment such as exist for the stackable frequency arrangement on pins 3 and 4, 9 and 10, 11 and 12, and 13 and 14.



LENGTH - LOSS CURVES			
FREQ.	TYPE OF SYSTEM	CHAN NO.	CARRIER
96KC	STACK'BL	4	UPPER
	NORMAL	4	UPPER

Fig. 4. A 96-kc maximum loss chart



LENGTH - LOSS CURVES			
FREQ	TYPE OF SYSTEM	CHAN NO.	CARRIER
36KC	STACK'BL	2	LOWER
	NORMAL	3	LOWER

LENGTH - LOSS CURVES			
FREQ	TYPE OF SYSTEM	CHAN NO.	CARRIER
30KC	STAGG	2	LOWER

LENGTH - LOSS CURVES			
FREQ	TYPE OF SYSTEM	CHAN NO	CARRIER
24 KC	STACK'BL	1	UPPER
	NORMAL	2	LOWER

LENGTH - LOSS CURVES			
FREQ.	TYPE OF SYSTEM	CHAN NO.	CARRIER
18KC	STAGG	1	LOWER

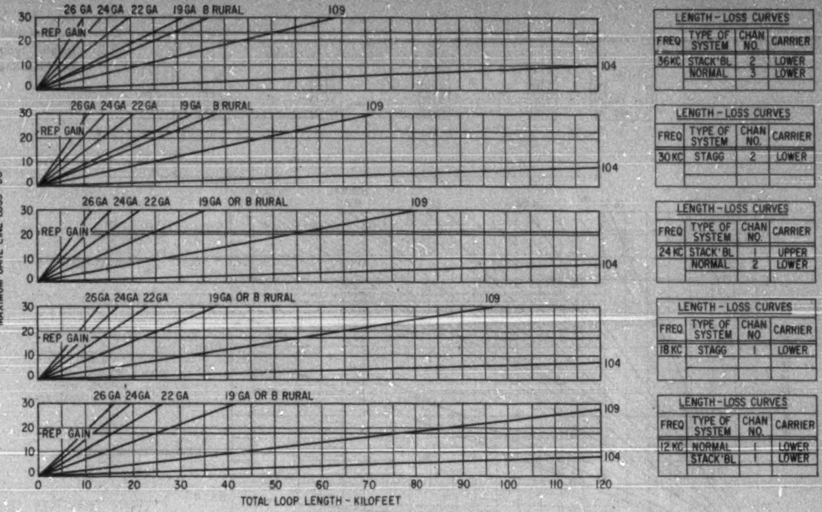
LENGTH - LOSS CURVES			
FREQ.	TYPE OF SYSTEM	CHAN NO.	CARRIER
12 KC	NORMAL	1	LOWER
	STACK'BL	1	LOWER

Fig. 5. Maximum loss charts

Length-Loss Carrier Line Design

To show how a carrier line design is arrived at by use of length-loss charts, the following example is presented to give the step-by-step procedures required. Fig. 7 shows in schematic form a typical rural layout indicating the gauge and type of existing facilities and the existing 1-, 5-, and 10-year line fills that will be encountered at major points along the route. Schematics of this type are the first step in engineering additional telephone facilities after a commercial forecast has been made to determine the potential demand for telephone service in the area to be served.

Referring to Fig. 7, the present facilities beyond the existing cable consist of 5 pairs of *109* steel wire, and the line fill requirements indicate that 7 additional pairs are needed initially with a total of 15 additional pairs in 10 years. Therefore, the problem is to determine the possibilities of carrier in serving these requirements versus other types of facilities such as cable, additional open wire,



LENGTH-LOSS CURVES

FREQ	TYPE OF SYSTEM	CHAN NO.	CARRIER
36KC	STACK'BL	2	LOWER
	NORMAL	3	LOWER

LENGTH-LOSS CURVES

FREQ	TYPE OF SYSTEM	CHAN NO.	CARRIER
30KC	STAGG	2	LOWER

LENGTH-LOSS CURVES

FREQ	TYPE OF SYSTEM	CHAN NO.	CARRIER
24KC	STACK'BL	1	UPPER
	NORMAL	2	LOWER

LENGTH-LOSS CURVES

FREQ	TYPE OF SYSTEM	CHAN NO	CARRIER
18KC	STAGG	1	LOWER

LENGTH-LOSS CURVES

FREQ	TYPE OF SYSTEM	CHAN NO.	CARRIER
12KC	NORMAL	1	LOWER
	STACK'BL	1	LOWER

Fig. 6. Minimum loss charts

or *B* rural wire. The use of carrier versus cable or open wire to care for the facility requirements of a particular job must, of course, include a cost study to determine how much carrier and how much physical plant should be added to provide the most economical arrangement. However, after experience is gained in a given area with the use of carrier, these studies will be simplified to the extent that a prove-in length of carrier versus physical plant can be established. Thus, the prove-in length can be used to determine the economical solution very easily. Therefore, in the following it is assumed that the proper economic solution has been determined and the use of carrier as applied provides the economical carrier line design.

Referring to Fig. 7, between 20 and 40 kilofeet (kf) only four circuits will be used in the 10-year period, and between 40 and 60 kf only three circuits are used in the 10-year period. Therefore, the existing VF facilities, with the addition of 20 kf of 6-pair *B* rural wire between 20 and 40 kf would provide all the necessary service for the 10-year period out to 60 kf. The increase of nine circuits beyond 60 kf of which six extend to 90 kf and three beyond 90 kf can be served by carrier. Therefore, the first step is to determine the number and location of repeaters required and the proper channel allocations to the existing facilities to provide a tentative carrier layout of the initial, 5-year, and 10-year additions. To do this, a plot of the appropriate length of each type of facility is made on the 96-kc length-loss chart, as shown in Fig. 8. It can be seen from this chart that the remote terminal of channel 4 could be placed at 97 kf with a maximum of three repeaters per carrier system. Therefore, there will be no problem in providing these circuits via carrier.

The next step is to determine the best method of carrier frequency arrangements so that the minimum number of carrier systems and, thus, the minimum number of repeaters will be required. An expanded sketch of Fig. 7, as shown in Fig. 9, can now be made to indicate the number of channels needed for the three periods of growth on which the remote terminals are designated in accordance with service requirements. This sketch indicates that 6 carrier channels will be needed initially, 9 carrier channels in 5 years, and 11 carrier channels at the end of the 10-year period. Therefore, two normal group frequency arrangements and one staggered frequency arrangement will fill the 10-year requirements. Referring to Table I for open-wire carrier system assignments, it can be seen that four normal frequency channels can be applied to pins 1 and 2, and 7 and 8, and the three staggered frequency channels can be assigned to pins 3 and 4 or 9 and 10. Therefore, the third step would be to make a pole-head schematic to show the existing pin positions of the wire presently in plant from 60 kf on out, from which could be determined the best carrier assignment, as shown in Fig. 10(A) (B), and (C). In Fig. 10(A), the initial carrier assignments are made on pins 3 and 4, and 7 and 8 with a complete system of four channels using the normal frequency arrangement placed on pins 7 and 8 and two channels of the staggered frequency arrangement operating on pins 3 and 4 in the multicircuit portion of the loop, and the actual terminals working from a bracket circuit extending beyond 90 kf.

Once the terminals have been located, the carrier design is made final by determining the values of input and output pads to keep the circuit net loss within the automatic regulation range of the equipment. As an example, the pad values for channel 1*S* are determined by plotting the facility loss on the 66-kc maximum loss chart and the 18-kc minimum loss chart. The origin of the chart represents the last repeater location or 63 kf in the example for the maximum loss chart and the terminal location for the minimum loss chart, since the higher frequencies are transmitted from the central office to the remote terminal and the lower frequencies conversely.

(*Continued on page 213*)

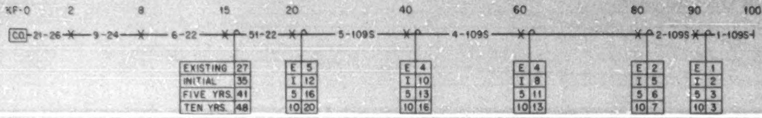

KF-0
2
8
15
20
40
60
80
90
100
C.O.
21-26
9-24
6-22
51-22
5-109S
4-109S
2-109S
1-109S
EXISTING 27
INITIAL 35
FIVE YRS. 41
TEN YRS. 48
E 5
I 12
5 16
10 20
E 4
I 10
5 13
10 16
E 4
I 8
5 11
10 13
E 2
I 5
5 6
10 7
E 1
I 2
5 3
10 3

Fig. 7. Schematic of rural line length type of facilities, line fills

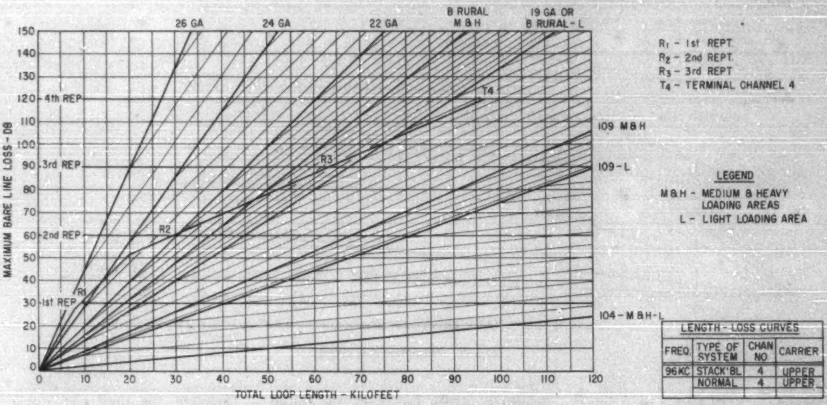


Fig. 8. Rural line layout on 96-kc maximum loss chart

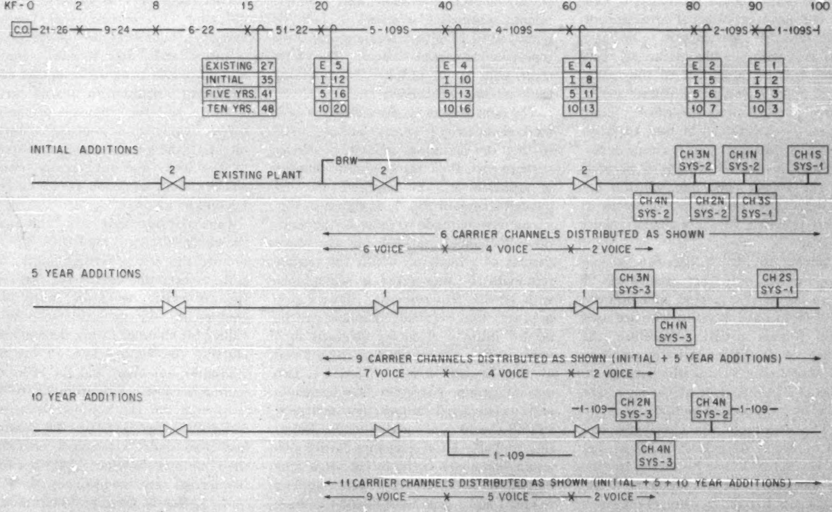

KF-0
2
8
15
20
40
60
80
90
100
C.O.
21-26
9-24
6-22
51-22
5-109S
4-109S
2-109S
1-109S
EXISTING 27
INITIAL 35
FIVE YRS. 41
TEN YRS. 48
INITIAL ADDITIONS
EXISTING PLANT
BRW
2
CH 3N SYS-2
CH 1N SYS-2
CH 1S SYS-1
CH 4N SYS-2
CH 2N SYS-2
CH 3S SYS-1
6 CARRIER CHANNELS DISTRIBUTED AS SHOWN
6 VOICE
4 VOICE
2 VOICE
5 YEAR ADDITIONS
1
CH 3N SYS-3
CH 2S SYS-1
CH 1N SYS-3
9 CARRIER CHANNELS DISTRIBUTED AS SHOWN (INITIAL + 5 YEAR ADDITIONS)
7 VOICE
4 VOICE
2 VOICE
10 YEAR ADDITIONS
1-109
CH 2N SYS-3
CH 4N SYS-2
CH 4N SYS-3
11 CARRIER CHANNELS DISTRIBUTED AS SHOWN (INITIAL + 5 + 10 YEAR ADDITIONS)
9 VOICE
5 VOICE
2 VOICE

Fig. 9. Schematic of carrier application

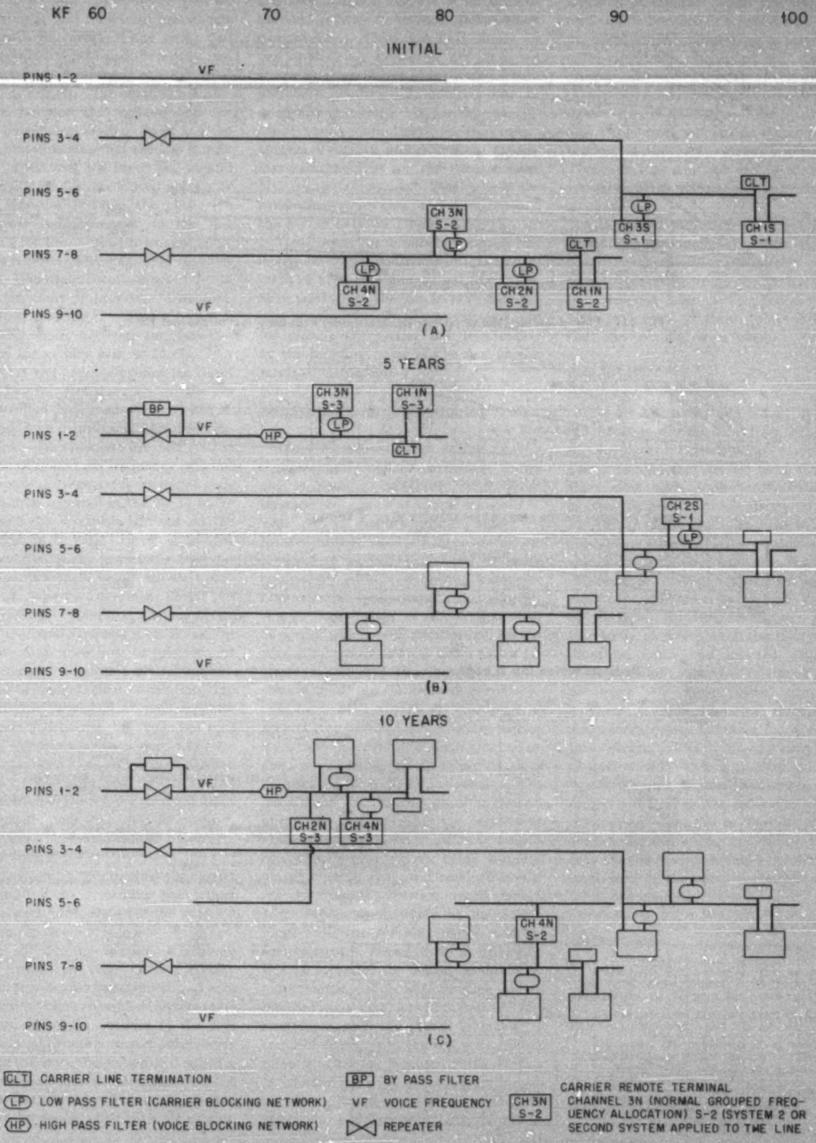

KF 60
70
80
90
100
INITIAL
PINS 1-2
VF
PINS 3-4
PINS 5-6
CLT
LP
CH 3S S-1
CH 1S S-1
CH 3N S-2
LP
PINS 7-8
CLT
LP
LP
CH 4N S-2
CH 2N S-2
CH 1N S-2
PINS 9-10
VF
(A)
5 YEARS
BP
CH 3N S-3
CH 1N S-3
PINS 1-2
VF
HP
LP
CLT
PINS 3-4
CH 2S S-1
LP
PINS 5-6
PINS 7-8
PINS 9-10
VF
(B)
10 YEARS
VF
HP
PINS 1-2
CH 2N S-3
CH 4N S-3
PINS 3-4
PINS 5-6
CH 4N S-2
PINS 7-8
PINS 9-10
VF
(C)
CLT CARRIER LINE TERMINATION
LP LOW PASS FILTER (CARRIER BLOCKING NETWORK)
HP HIGH PASS FILTER (VOICE BLOCKING NETWORK)
BP BY PASS FILTER
VF VOICE FREQUENCY
REPEATER
CH 3N S-2 CARRIER REMOTE TERMINAL CHANNEL 3N (NORMAL GROUPED FREQUENCY ALLOCATION) S-2 (SYSTEM 2 OR SECOND SYSTEM APPLIED TO THE LINE

Fig. 10. Expanded view of rural line schematic showing open-wire carrier layout

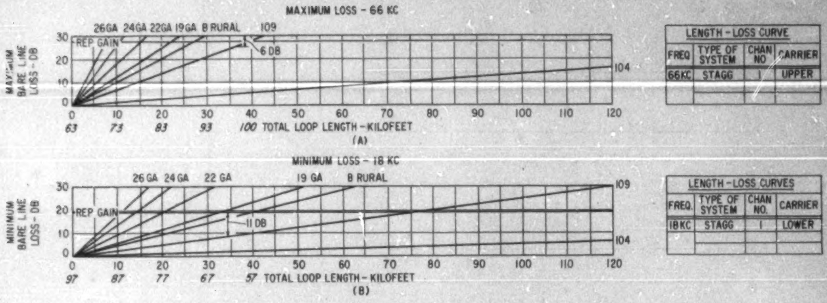


Fig. 11. Plot of last repeater to remote terminal carrier frequency loss of channel 1S

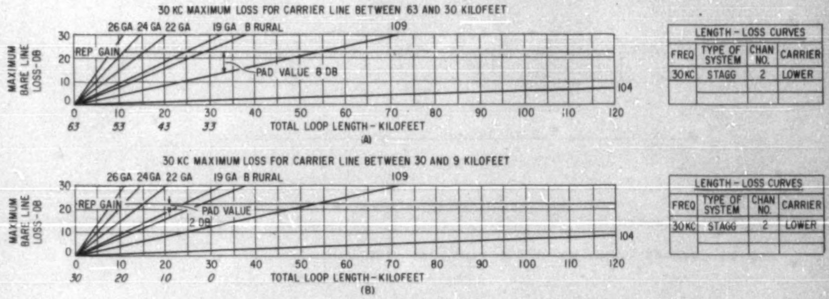
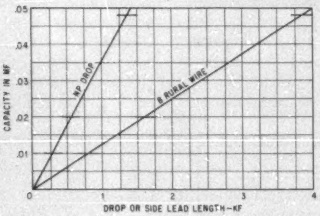


Fig. 12. Repeater section loss at mid-band carrier frequency in remote-to-central-office direction



NOTE:
WHEN TOTAL CAPACITY OF SIDE LEAD DOES NOT EXCEED:-
0.01MF-USE 482A NETWORK AT JUNCTION
0.02MF-USE 482B NETWORK " "
0.048MF-USE 482C NETWORK" "
WHEN TOTAL CAPACITY EXCEEDS 0.048MF USE 482B AT JUNCTION AND 482D AT END OF LEAD

USE OF CHART: APPLICABLE TO EITHER SUBSCRIBER DROPS OR VF SIDE LEAD CIRCUITS CONNECTED TO A PAIR ON WHICH CARRIER FREQUENCIES ARE PRESENT AT SUCH BRIDGING POINTS.

A. FOR EACH LENGTH OF EACH FACILITY IN THE LOOP FROM THE MAIN CARRIER LINE OBTAIN THE APPROPRIATE VALUE OF CAPACITY FROM THE CHARTS.

B. COMPUTE THE TOTAL CAPACITY AND DETERMINE WHICH OF THE ABOVE NETWORK ARRANGEMENTS SHOULD BE MADE.

EXAMPLE: ASSUME SIDE LEAD COMPOSED OF 1000 FT. OF B RURAL WIRE, 6000 FT. OF 109S LINE WIRE AND 200 FT. OF NP DROP.

FACILITY	LENGTH	CAPACITY
B RURAL	1000 FT.	.0126 MF
109	6000 FT.	.0105 MF
NP DROP	200 FT.	.0072 MF
	TOTAL	.0303 MF

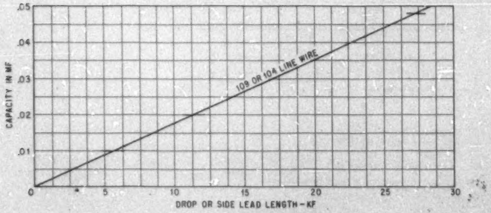


Fig. 13. Filter distribution length—capacitance curves for the selection of carrier-blocking network

Vol. 7
Pt. 1
1956

(Continued from page 209)

Figs. 11(A) and (B) show the plot on the maximum and minimum loss charts from which the value of the input pad at the remote terminal and the output pad at the remote terminal can be read directly as follows: the maximum allowable loss between the repeater and the remote terminal is 30 db at 66 kc and the facility plot indicates that the loss is only 24 db. Therefore, the input pad should be the difference in these two values, or 6 db. Fig. 11(B) shows the plot from the remote terminal back to the last repeaters for the minimum loss of 18 kc which is 8.5 db. The maximum repeater gain at this frequency, shown by the line on the chart, is 19.5 db. Therefore, the output pad for the remote terminal is the difference between the repeater gain and the facility loss, or 11 db.

The next step is to determine the input pad values for the repeaters in the remote-terminal-to-central-office direction. These pads are required in this direction only, since the repeaters are laid out on the basis of maximum frequency and the gain slope of the repeater is such that pads would not be required in the 60- to 96-kc band with the repeaters laid out on the basis of a 96-kc loss. However, in the 12- to 48-kc band, the variation in repeater gain is sufficient to result in the possibility of repeater overload based on 96-kc spacing. Therefore, the maximum loss at midband frequency, or 30 kc for the low-group direction, is plotted for each repeater section in the remote-terminal-to-central-office direction, and an input pad is placed at each repeater which is equal to the difference between the maximum facility loss and repeater gain at 30 kc. This is illustrated in Figs. 12(A) and (B). In a similar manner, each of the remote terminal input and output pads can be determined, thus giving a complete carrier line transmission design.

The next step in completing the carrier line design is to determine the proper filter network to be placed at the junction of the carrier line and side leads, or drops, serving VF derived circuits, which are served from the base carrier facility between remote carrier terminals. These filters are required to reduce the reflection losses at carrier frequencies of the side leads and to prohibit carrier from being transmitted into the subscriber subsets, since certain types of subscriber subsets, such as the *500* set which is the recommended set to be used with *P1* carrier, are susceptible to carrier frequencies from a noise standpoint because of the demodulation qualities of certain of the subset components. These low-pass filters or carrier-blocking networks are designed to match given ranges of capacitance that will be encountered from drop wire or open-wire side leads. Since this capacitance varies considerably with various lengths of facility, and equivalent capacitance-length chart for the various types of facility is provided from which the total capacitance can be determined and the proper network chosen. See Fig. 13.

This covers the details of a carrier line design with the exception of the placement of high-pass filters used to sectionalize portions of the line for VF distribution where filter distribution is used and for placing the carrier line termination at the bridging point of the last carrier terminal on the line.

Installation and Maintenance

The *P1* carrier system utilizes plug-in equipment throughout primarily as an aid in making installation and maintenance procedures as simple as possible. The plug-in feature allows individual printed wiring boards or an entire carrier terminal to be easily replaced by spare equipment without requiring extensive maintenance procedures. To simplify this process further, a test set has been designed which will provide all of the necessary carrier and VF and a means for measuring them. The installation consists of lining up the over-all carrier circuit for specified values to be read at special test points built into the equipment, and of adjusting certain potentiometers or inserting given values of pads to meet the over-all requirements. The installation and line-up procedure is set up on the basis of using the test set to make a series of measurements in a specified order, with signals being supplied at the central office and received at the remote terminal or vice versa and of adjusting potentiometers in the equipment until the specified meter readings are obtained. The original line-up results will be recorded for future reference on maintenance calls. Maintenance will be handled on a complete terminal replacement basis and will consist of making a series of checks with the test set to determine whether the terminal is functioning satisfactorily. If it is not, and if it cannot be adjusted to restore satisfactory operation, a replacement terminal will be used to restore service. All repairs and isolation of trouble within the terminal unit will be handled at a centralized testing point so that a minimum of personnel with electronic experience will be required. However, the test set itself has been designed to handle all tests for a *P1* carrier system and, when used by experienced personnel, will enable the complete testing of the *P1* carrier equipment. Therefore, one test set can be used for installation, maintenance, and trouble isolation.

The remote terminals are hermetically sealed and test points have been brought out to the terminal strip outside of the remote terminal housing so that the housings will not have to be opened unless tests performed at these points indicate that the terminal is in trouble.

Conclusion

The use of carrier equipment in large quantities to provide subscriber telephone service places requirements on equipment and personnel that are far more difficult than exist with equipment located in central offices or fixed places. Therefore, every effort has to be made to make the maintenance of this equipment as simple as possible. Also, the flexibility required to provide subscriber service makes it necessary to use a wide variety of facilities in the exchange plant. Carrier is merely one of the many that must be considered by the engineer. Thus, for the proper use of carrier, it is necessary to have some simplified methods of application that allow a fast and accurate way of laying out a carrier line design. The use of length-loss charts provide such a method and the use of plug-in equipment with designated test points and a test set with detailed instructions allow for an over-all ease of engineering and application that is necessary for the present technical ability and work load of the operating forces of the telephone company.

References

1. Objectives and General Description of Type-P1 Carrier System, R. C. Boyd. *AIEE Transactions*, vol. 75, pt. I, May 1956 pp. 188–91.

2. Power Supplies for the Type-P1 Carrier System, D. H. Smith. *Ibid.*, pp. 191–95.

3. Simplified Transmission Engineering in Exchange Cable Plant Design, L. B. Bogan, K. D. Young. *Ibid.*, vol. 73, pt. I, Nov. 1954, pp. 498–502.

4. Telephone Lines for Rural Subscriber Service, Lester Hochgraf, R. G. Watling. *Ibid.*, vol. 74, pt. I, May 1955, pp. 171–76.

Discussion follows

Discussion

J. M. Moss and **D. I. Cone** (The Pacific Telephone and Telegraph Company, San Francisco, Calif.): The application methods described by the author are a far cry from those used in toll line carrier design. Great simplification is possible because of the basic design of the system and the use to which it will be put. Even so, it will mean a great extension of the efforts of the engineer normally engaged in the design of facilities for rural service, beyond his previous customs. Operation and maintenance will likewise change.

Thus it appears that a considerable program of instruction will be required to acquaint these people with the layout procedures, including a sufficient appreciation of the characteristics of carrier transmission to enable them to recognize the need for control of reflections by impedance-matching, line terminations, and filter placement. This same problem faced the central office people when electronic gear first appeared on toll circuits. Questions that seemed immensely complex at first are now solved in routine fashion, and there is every reason to believe that the same result will be achieved with the local plant.

Telephone engineers throughout the country will surely welcome the advent of the *P1* carrier as a powerful aid in expanding rural telephone service. In the problem of determining the economical means of service, devices such as the chart in Fig. 8, serve to make possible quick analysis of the conductor and repeater requirements covering the numerous gauges of cable and open-wire types. Thanks to the work of the research and development groups, standards are set up to make possible a sort of "mass production" application. The burden of determining the standards of design has been recognized and resolved in lengthy interchange of information and ideas between operating people in the field and the group of designers whose conclusions are embodied in this paper and its companions.

Having arrived at this standard, the next step is its application in the field and the inevitable experiment to determine what variations can be profitably made. Those who have experienced the development in application of earlier carrier systems will recognize the possibilities. There is a genuine engineering challenge both to find variations of application that are useful and to understand the fundamental limitations of each such carrier system.

J. M. Flanigan (U. S. Department of Agriculture, Washington, D. C.): The paper gives an excellent presentation of the manner in which the type-*P1* carrier system can be applied to rural subscriber loops. The method of using loss charts for determining repeater and terminal spacing is particularly interesting. In the application of subscriber carrier systems of various types for rural telephone use, however, we have found that numerous situations occur where it is necessary to utilize the same central office entrance cables and portions of the open-wire routes that are used for existing or proposed trunk carrier systems. For these applications the problem of providing maximum co-ordination between various types of carrier systems becomes more difficult, and consideration must be given to the operating frequencies, levels, crosstalk loss achieved between the various wire facilities upon which carrier channels are operated, and the susceptibility of various types of carrier equipment to interfering effects. To achieve maximum economies from the use of subscriber carrier systems in many application situations, co-ordination between carrier systems must definitely be considered initially in the layout of the carrier routes.

There are no illustrations which show the central office carrier terminals mounted at locations external to the central office, such as at cable open-wire junctions. In this type of arrangement more cable pairs are required for the transmission of voice frequencies; there is a maximum distance between the central office and an externally mounted central office carrier terminal over which the desired ringing information can be properly transmitted; and carrier channel testing cannot conveniently begin at the central office. Despite these apparent disadvantages, application situations might arise whereby the placement of a carrier terminal at an external location not exceeding the maximum distance mentioned may reduce the over-all cost of providing service by subscriber carrier techniques by reducing the number of carrier repeaters required in the over-all system arrangement. Although testing of a carrier channel cannot begin in a central office with this arrangement the over-all testing time might still be reduced if fewer repeaters are required.

The paper gives considerable detail as to how to calculate losses for both open-wire and cable facilities at carrier frequencies. Little detail is given, however, as to how the effective transmission objective is altered for the subscriber loop, portions of which are also utilized for carrier transmission, by the insertion of low-pass filters at various locations along this loop. These filters provide an insertion loss at voice frequencies since at least two of them are inserted in the loop on a call between a voice frequency subscriber and the central office. In addition, the remaining low-pass filters would be bridged on the loop in an unterminated condition during a call and would probably cause additional transmission losses at voice frequencies.

Since commercial power for the low-drain chargeable battery type of power supply will be the most economical for powering each remote carrier channel terminal, a minimum power bill will probably have to be paid by the telephone company for each separate remote terminal location despite the very low current drain of the transistorized equipment. If remote terminal locations were combined by the economical utilization of various arrangements of filter distribution, the costs for minimum power bills would be minimized. In addition, a minimum number of remote carrier terminal locations would reduce the time required for periodically testing and maintaining this equipment. If a minimum number of remote terminal locations were economical, an a-c powered low-drain chargeable battery type of power supply with sufficient capacity for powering more than one remote terminal could possibly be more advantageous from an equipment cost standpoint as compared with a separate power supply for each channel terminal, as now provided.

J. D. Howard, Jr.: The discussions are interesting and important. I agree with Mr. Moss and Mr. Cone in regard to the necessity of more extensive training programs for acquainting the Bell System plant engineers and operating people in the concepts of carrier. Also, the possibilities of further techniques in carrier application should certainly stimulate the engineer's thinking and challenge his imagination and intellectual curiosity.

Mr. Flanigan has pointed out several of the engineering aspects of carrier application that need clarification. For example, the possible co-ordination of *P1* carrier and other types of carrier systems. It will certainly be necessary to consider alternate frequency assignments, crosstalk computations, and level differences to determine the optimum carrier application. These techniques are not new in the telephone industry and, rather than attempting to provide training to all of the plant engineers, the company will of necessity have the techniques performed by specialists in this type of work.

As to remote location of the central office terminal of *P1* carrier it should be pointed out that this was not considered in the original equipment design as a requirement because of the decision to provide a carrier system operable over both cable and open wire with repeaters to counteract the high loss of cable plant at carrier frequencies. However, if this becomes a necessity, further refinement of the system might possibly be made, but it must be remembered that this means an additional option and a greater cost that might well outweigh the advantage, since low cost of *P* carrier is of paramount importance. The placement of the central office terminal at a remote location may also require the use of VF repeaters on all of the carrier circuits at the central office to make up for the VF loss of the entrance cable to ensure the proper grade of transmission.

The effective loss of a carrier channel and its VF extension depends on the following factors:

1. Effective transmission of the VF extension.
2. Circuit net loss of the carrier channel and possible variations in this loss.
3. Side-tone impairment due to the use of carrier in the loop.
4. Impairments due to the use of *482*-type networks (low- and high-pass filters).

It is estimated that the total impairment of these filters will be in the order of ±3 db that would be added to the over-all effective loss of a VF rural loop with the exact value depending upon the possible combination of factors involved in a particular situation.

Mr. Flanigan's idea on conserving commercial power taps by combined locations of terminals is very well taken, and the Bell Telephone Laboratories are presently considering the possibility of providing a power unit for more than one terminal. A single battery supply for more than one terminal presents a problem of battery supply filtering to eliminate the possibility of crosstalk between channels, and it may prove to be uneconomical, especially considering that a multiunit power supply may well be a low-demand item, thus further increasing production costs.

Co-ordinated Protection for Open-Wire Joint Use—Present Trends

W. R. BULLARD
FELLOW AIEE

H. E. WEPPLER
MEMBER AIEE

THE EXTENT of use and benefits obtained by the joint use of pole lines to carry power and communications circuits are well known. A number of worthy papers have been presented before the Institute to describe various aspects of joint use and to report on progress in this field.

The power companies and communication companies have long worked in close co-operation to develop techniques which would make possible even more extensive realization of the utility and economy of joint use while still safeguarding the public, the employees, and the essential public services rendered.

As the power and communication systems are extended and adapted to serve best the public's needs, new problems in joint use constantly arise. The extension of joint use from a predominately urban field of application to widespread use in rural areas was an important step which has previously been reported.[1] Similarly, considerations involved in joint installations with power lines of increased operating voltages have been presented.[2,3]

This paper reviews some of the problems encountered in the past few years. Two other papers[4,5] describe field tests which are leading the way to solutions of current problems.

Objectives for Protection

The probability of contact between power lines and telephone lines varies, of course, with the types of plant involved. For example, with power conductors in cable the chance of a failure resulting in a contact with the telephone conductors is remote. However, for obvious economic reasons large portions of the power and telephone plant must utilize aerial wire construction. In joint-use construction of this type, the possibility of contact assumes greater importance. The prime factors in holding to a minimum the occurrence of contacts between the power and telephone conductors are the development and consistent use of safe structural standards. The practices in use have been developed co-operatively over the many years of joint-use construction and are frequently reviewed in the light of current needs.

In the infrequent case where contact does occur the objectives are to minimize the foreign potential developed on the telephone plant and to remove the power voltage from the telephone plant as quickly as possible.

The voltage impressed on the telephone plant during contact is determined by the power circuit voltage and the impedance of the telephone plant between the point of contact and ground as compared with the total impedance of the fault circuit. In the telephone plant it is thus essential to provide for low impedances to ground during faults in order to hold down the foreign potentials.

The length of time the telephone plant is energized is determined in most cases by the time–current characteristics of the interrupting devices in the power circuit and by the magnitude of the fault current. There are, of course, a fraction of the cases in which the contact is only momentary due to the physical nature of the contact itself, and there may be some in which the contact is interrupted by the burning away of the telephone conductors or power conductors at the point of contact. However, electrical protection on joint-use lines is primarily a matter of co-ordinating the ratings of power circuit interrupting devices, power system impedance, telephone system impedance, and telephone system protective devices.

Effect of Growth in Power Loads

With the increasing power loads of recent years, the power companies in a number of cases have found it necessary to increase line voltages to increase the capacity of their lines. At first the joint use of pole lines with power voltages over 5 kv was looked upon as something quite special, but subsequent study and experience have established that properly co-ordinated joint use need not be limited to any arbitrary voltage.[3] In fact, the use of higher power voltages has in many cases facilitated the rapid de-energization of the power line since the ratio of fault current to load current is larger with the higher line voltage.

Still more recently, however, it has become necessary to go to progressively higher currents in some heavy-duty distribution systems. In such instances the impedance of the power circuit is correspondingly lower because of the heavier lines and equipment required to carry the greater load, and because of the influence on design of voltage-drop considerations in normal operation. To minimize the foreign potential on telephone plant in the event of contact, and to insure prompt de-energization, it then becomes necessary for the telephone plant to provide lower impedance paths to ground than heretofore and to handle increased fault currents.

In some other cases where load capacity is increased by voltage increases with no significant change in load current, the possible use of fault-current limiting devices may create a similar need for low impedance in the telephone plant in order to insure fault-current values which can be distinguished from load currents and initiate prompt de-energization of the power line. The problem of handling a wide range of fault currents in telephone plant is complicated further in some cases by the use of power system reclosers with fast–slow sequences to permit automatic sectionalization of the power loads. Such operation, of course, extends over a longer period of time than in the cases where positive and rapid single-shot interruption is provided.

Telephone System Protection

In the telephone plant, carbon block arresters which provide a low-voltage spark gap are used at subscriber premises to provide effective grounding of the telephone lines when the lines are subjected to abnormal voltages.[6] In order to minimize potential differences which might constitute fire or shock hazards, the grounding is preferably common with that used by the power system and is, of course, as low in impedance as possible.

At telephone stations served from open wire the arresters are protected against overheating by 7-ampere fuses which, in the case of very high currents or sustained currents, open and isolate the station and

---

Paper 56-237, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing December 6, 1955.

W. R. Bullard is with Ebasco Services, Inc., New York, N. Y., and H. E. Weppler is with American Telephone and Telegraph Company, New York, N. Y.

the station protector to prevent overheating. In the case of stations served from metallic-sheathed cable, a fuseless type of station protector is used which is capable of withstanding higher currents and provides a continuous path to ground in the case of foreign voltages.

Where distribution is by means of metallic-sheathed cable, the cable sheath is grounded frequently in order to keep the impedance to ground low. Where joint with multigrounded neutral power lines, the sheath is often bonded to the power system neutral at frequent intervals to provide a low-impedance path for fault currents.

In the open-wire plant additional heavy-duty carbon block arresters are used where the plant is exposed to power lines operating at 5 kv or higher voltages in order to provide supplementary paths to ground in the event of contact. These operate after the station arresters and fuses have operated, grounding the telephone line to prevent arc over of station fuses and to provide prompt and positive de-energization of the power circuit. The protector used for many years in this application is the Western Electric *99A* protector (or other similar types) which operates to ground the telephone conductors when the impressed voltage exceeds about 3,000 volts.

Minneapolis Tests

Because of the prospect of heavier suburban power loads in their area, with the attendant problem of co-ordinating protection, the Northern States Power Company, the Northwestern Bell Telephone Company, and the Bell Telephone Laboratories in 1953 conducted a series of field tests of protector arrangements at a site near Minneapolis, Minn. This work is described in a paper by Albrecht and others.[4]

The data desired included an extension of the time–current curves for the *99A* protector into the range of higher currents than previously tested and the effects of various recloser cycles. The tests provided valuable information on the limitations of the *99A* and offered the opportunity for initial tests on an experimental protector design proposed for the heavier fault currents and for the extended duty cycles which might be expected in the future.

The Minneapolis tests also afforded an opportunity to investigate the fusing or burn-down characteristics of the common types of telephone line wires. Such characteristics become of particular significance as heavy-duty protector designs approach the point where the protector current-handling capacity may exceed the fusing points of line conductors. It was first recognized as a result of these tests that the change in impedance of telephone conductors while carrying fault currents may have a significant effect on the co-ordinated protection.

Ontario Tests

Subsequent to the Minneapolis tests, the Bell Telephone Laboratories proceeded with the development of a new heavy-duty protector, the Western Electric *118A*, along the lines of the experimental type tried at Minneapolis.

Preproduction samples of the *118A* were tested in 1954 during a series of co-operative tests conducted near Toronto, Ont., Canada, by The Hydro-Electric Power Commission of Ontario, The Bell Telephone Company of Canada, and the Bell Telephone Laboratories. That work is reported in detail in the paper by Ellis and others.[5]

The Ontario tests provided valuable information on the time–current characteristics of the *118A* protector. A limited number of tests on the *99A* were included also, to provide direct correlation between the two types. Additional data on fusing and change-in-impedance characteristics of telephone line wires and cable were also obtained to augment the earlier tests.

By means of this series of tests it was determined that the current-carrying capacity of the *118A* protector will be much greater than that of the *99A* and will exceed the fusing characteristics of the more commonly used exchange telephone line wires. Several refinements for the *118A* design were suggested as a result of the tests.

Conclusions

The protector development work and field tests of the past 2 years have resulted in a new protector design which is expected to make possible co-ordinated protection of telephone lines and high-capacity power lines, thus facilitating joint use of pole lines for such facilities.

The new protector is expected to have current-handling properties exceeding those of the most commonly used exchange telephone conductors, suggesting that in those cases any further increase in protector capacity would be unnecessary.

As the ratio of estimated fault current to load current is reduced in typical cases, it will be increasingly important to take into account the thermal characteristics of the telephone conductors in estimating the impedance of fault-current paths. There is a possibility that the thermal rise in line impedance may cause the fault current to drop below that which would insure positive de-energization of the power line.

The possibility is foreseen of cases in which the fusing characteristics of the telephone line wire would be exceeded by the estimated fault currents and in which it might be found desirable to take account of the burn-down of the telephone conductors to provide the means of removing the foreign potential from the telephone plant rather than to rely on the de-energization of the power line itself.

References

1. Joint Use of Pole Lines for Rural Power and Telephone Services, J. W. Campbell, L. W. Hill, L. M. Moore, H. J. Scholz. *AIEE Transactions*, vol. 66, 1947, pp. 519–24.

2. Protection Features for the Joint Use of Wood Poles, J. O'R. Coleman, A. H. Schirmer. *Ibid.* (*Electrical Engineering*), vol. 57, Mar. 1938, pp. 131–40.

3. Some Aspects of Joint Use of Wood Pole Lines with Increasing Line Voltage, W. E. Bloecker, W. R. Bullard. *AIEE Transactions*, vol. 72, pt. I, 1953, pp. 709–12.

4. Co-ordinated Protection for Open-Wire Joint Use—Minneapolis Tests, E. G. Albrecht, E. W. Christoferson, A. E. Dietz, J. C. Slothower. *Ibid.*, vol. 75, pt. I, May 1956, pp. 217–23.

5. Co-ordinated Protection for Open-Wire Joint Use—Ontario Tests, H. M. Ellis, J. W. Phelps, C. L. Roach, R. E. Treen. *Ibid.*, pp. 223–36.

6. Electrical Protection of Telephone Systems, W. R. Bullard, J. B. Hays, H. O. Saunders. *Ibid.*, vol. 73, pt. I, 1954, pp. 385–88.

Co-ordinated Protection for Open-Wire Joint Use—Minneapolis Tests

E. G. ALBRECHT
NONMEMBER AIEE

A. E. DIETZ
NONMEMBER AIEE

E. W. CHRISTOFERSON
NONMEMBER AIEE

J. C. SLOTHOWER
ASSOCIATE MEMBER AIEE

THE rapid growth of suburban areas in the vicinity of the larger cities has created many problems for both power and communication companies. A number of power companies have chosen to meet load growth in such areas by increasing their primary distribution voltage. Such increases in voltage have in general been accompanied by substantial increases in substation, feeder, and subtransmission capacity. These changes have in turn introduced problems concerning the maintenance and operation of overhead lines jointly occupied by the power and communication companies.

The Northern States Power Company was faced with such load growth problems following World War II in the suburban areas adjacent to Minneapolis and St. Paul, Minn. The solutions to these problems had to take into consideration the extensive joint use of wood poles with the Northwestern Bell Telephone Company in these areas.

As a result of an extensive and detailed engineering study, the decision was made to convert most of the existing 4-kv multigrounded neutral system in these areas to 12.5 kv. The conversion, which was to take place over a period of 5 to 7 years, was complicated by the fact that a great deal of open-wire telephone plant was involved. The extensive use of exchange open-wire telephone plant in these areas, most of which was on joint-use poles, was due to the relatively low population density and the extensive use of multiparty service. Economic considerations precluded replacement of much of this open-wire plant with cable in conjunction with the conversion of the power system to 12.5 kv.

The generally accepted standards of good engineering as set forth in the National Electrical Safety Code[1] provide that if telephone circuits are to be placed on the same poles or involved in numerous crossings with power circuits at voltages above 5,000 volts, a higher grade of construction should be used for the power line except where co-ordinated electrical protection is provided.

Co-ordinated protection involves the application of special protective devices to the power and telephone systems which will permit both of the following conditions to be fulfilled:

1. The power and telephone circuits are so constructed, operated, and maintained that the power circuits will be promptly de-energized, both initially and following subsequent breaker operations, in the event of a contact with the telephone plant.
2. The voltage and current impressed on the telephone plant in the event of a contact with the power conductors are not in excess of the safe operating limits of the telephone protective devices.

In view of the impracticability of providing separate lines or providing a higher grade of construction in the proposed conversion areas, it was mutually agreed that the application of co-ordinated electrical protection was the proper engineering solution.

Joint occupancy of pole lines by telephone and power companies at voltages higher than 5,000 volts is not new.[2] Practically all of the previous joint-use experience with high-capacity higher voltage distribution systems, however, involved telephone cable plant rather than open wire. Previous experience with open-wire joint use with higher voltage power systems on the other hand has been almost entirely limited to low-capacity rural-type distribution systems.[3-6]

The development in recent years of new protective devices for power systems made possible the development of a heavy-duty high-capacity 12.5-kv distribution system by the Northern States Power Company.

The demand for information concerning the co-ordination of existing telephone protective equipment with these new power system protective devices resulted in the tests discussed in this paper. These tests were conducted jointly by the Northwestern Bell Telephone Company, the Northern States Power Company, and the Bell Telephone Laboratories at Minneapolis in 1953.

Power System Characteristics

The design features of the high-capacity heavy-duty 12.5-kv distribution system, developed by the Northern States Power Company, which affect protection co-ordination are as follows:

1. The maximum permissible symmetrical fault current for a zero-impedance ground fault is limited to approximately 4,000 amperes rms for operating reasons, and the usual minimum value for a similar fault is 600 amperes. Typical fault currents for zero-impedance faults at any point on a 12.5-kv feeder main line or tap as a function of distance from the source are shown in Fig. 1. The intersection of each numbered curve with the vertical axis represents the fault current on the feeder main line at that number of miles from the substation. Fig. 2 shows the typical fault currents for 20-ohm ground faults. Curves *A* and *B* of Fig. 2 are shown on Fig. 1 to illustrate the controlling effect of fault impedance.
2. Substations are spaced from 6 to 8 miles apart with from four to six 300-ampere feeders from each substation. Fig. 3 shows a typical 12.5-kv area.
3. Each feeder main line is limited to a length of 4 miles, with full 300-ampere capacity over the whole length. Full-capacity emergency switching is provided between adjacent feeders (Fig. 3).
4. No ground fault relaying is provided at the substation at this time. This type of relaying is now under consideration but its effect on joint-use co-ordination is not expected to be significant in this situation.
5. The use of automatic circuit reclosers with dual operating characteristics (the so-called "fast" and "slow" operations) in feeder branches, or in some cases in the main lines, provides automatic sectionalizing of the feeders but imposes a heavier burden on the telephone protection system than would result from the use of fuses. This heavier burden is the result of delayed opening and up to three circuit reclosings.
6. The use of fuses instead of reclosers would preclude large-capacity feeders and in some cases would have made voltage conversion impracticable. Fuses are used, however, in short single-phase laterals where use of a recloser or sectionalizer is not justified or where open-wire joint use alone may require them.
7. The maximum demand on the single-phase laterals is generally limited to 300 kva (about 40 amperes) to permit restora-

---

Paper 56-239, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing December 6, 1955.

E. G. ALBRECHT is with the Northwestern Bell Telephone Company, Minneapolis, Minn.; A. E. DIETZ is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.; E. W. CHRISTOFERSON and J. C. SLOTHOWER are with the Northern States Power Company, Minneapolis, Minn.

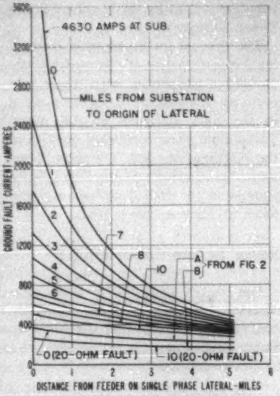


Fig. 1. Ground-fault currents for zero-impedance faults on a 12.5-kv feeder, 4/0 copper 3-phase lines with no. 6 copper single-phase laterals

tion of service with a 70-ampere recloser after loss of diversity. Selection of this recloser rating was influenced by the assumption of a 20-ohm maximum fault resistance in telephone contacts.

8. The assumption of a 20-ohm ground fault resistance results in current values which are close to main-line load currents. The phase relaying used on the feeders will not discriminate between ground fault and normal load currents. Ground fault relaying has been avoided up to the present because it is desirable to permit the feeders to operate with large unbalanced loads during emergency conditions.

Telephone System Characteristics

The voltage and current limitations of subscriber station protectors require the application of heavy-duty auxiliary protectors on open-wire exchange telephone circuits exposed to power system voltages in excess of 3,000 volts to ground. The Western Electric *99A* protector shown in Fig. 4, which is normally used in the Bell System for this purpose, consists of three cylindrical carbon electrodes enclosed in a weatherproof container and separated so that the breakdown voltage between them is about 3,000 volts rms. Two of the electrodes are connected to the telephone line wires and the third electrode is connected to ground. This must be a very low resistance ground which generally can be secured by bonding to an effectively grounded power system neutral.

These auxiliary protectors are normally installed at approximately half-mile intervals so that in the event of a contact midway between two protectors, assuming that only one protector operates, the fault impedance will not exceed 20 ohms. This value assumes a contact with the smallest size telephone conductor used on open-wire lines, *083* steel, although this size is infrequently used on joint-use lines. The fault impedance is reduced as the point of contact approaches the protector location and is zero at the protector. Consequently, for a given power system impedance the voltage on the telephone plant is proportional to the distance between the point of contact and the protector. The current, however, increases approximately inversely as this distance and may vary from 360 to 4,000 amperes for a power-telephone contact on the main line of a 12.5-kv feeder.

The current-carrying capability of the *99A* protector, published some 15 years ago, for three successive shots at 15-second intervals ranges from 600 amperes of 1-second duration to 1,800 amperes of 0.15-second duration. This was more than adequate for the open-wire joint use situations in effect at that time as the power system protection in most cases consisted of low-rated single or 3-shot fuses.

Test Objectives

As stated in the foregoing, a series of tests was scheduled for the summer of 1953 to secure the following information relative to power and telephone system protective devices which was not then available:

1. The time–current characteristics of the *99A* protector at higher current values.
2. The behavior of the *99A* protector under the duty cycles imposed upon it by various reclosers under a wide range of fault currents.
3. The capabilities of a number of other telephone protectors including several modifications of the standard *99A* and an experimental model of a new protector furnished by Bell Laboratories.
4. Data on the fusing, arcing, burn-down, and thermal characteristics of the types of telephone line wire most generally used in exchange plant.

Test Setup

The site selected for the tests was the Northern States Power Company's Gleasons Lake Substation. This substation is located about 10 miles west ofMinneapolis and is supplied from a direct tap on the 115-kv loop around Minneapolis and St. Paul. The substation was at that time a 7,500-kva 115/13.8-kv station with one express 13.8-kv tie-line to a near-by 13.8/4-kv substation and one 12.5-kv distribution feeder. Line protection consisted of station oil circuit breakers on the 13.8-kv line and the 12.5-kv feeder. Station protection consisted of a differential relay operating a high-speed grounding switch on the 115-kv side of the transformer.

The site itself consisted of several acres of ground more or less isolated from residential areas and free of trees and structures. A direct tap was made on one phase of the 13.8-kv line (which was removed from normal service for the tests) to an H-frame structure on which were mounted six reclosers, a potential transformer, and a hook-stick-operated cutout (see Fig. 5). From this structure one span of 336.4-MCM bare aluminum wire was strung at right angles to a span of telephone conductor and some 8 feet above it. I nthe actual tests a weighted contact wire from the power circuit was allowed to lie across the telephone wire or a direct connection was made from the power wire to the telephone wire. A 1-line diagram of the test layout is shown in Fig. 6.

The magnitude of the fault current was controlled by series-parallel combinations of three 100-foot spans of 48 wires each of telephone conductor, the particular wire used was *134H* steel wire. Normally, open-wire joint-use telephone plant in these suburban areas would consist of *109H* steel wire but the larger size was used in this case because of its greater thermal capacity.

In order to avoid dangerous earth potentials due to fault currents returning from the grid to the substation grounding system through the earth, one end of the grid was connected through an insulated cable directly to the substation ground bus. When telephone protectors were being tested they were inserted between the grid and the ground-return cable.

Oscillographic recordings were made of the current in the circuit and the voltage on the "load" side of the recloser with a 6-element oscillograph and with a 3-element rapid-record oscillograph. A timing wave taken from the station bus was also recorded on the 6-element instrument for reference purposes.

Test Results

The test results indicated that the *99A* protector currently available for open-wire joint use with higher voltage lines is inadequate for 12.5-kv high-capacity distribution systems similar to that de-

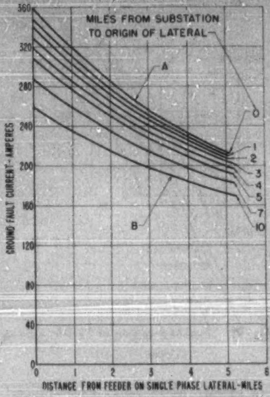


Fig. 2. Ground-fault currents for 20-ohm faults on a 12.5-kv feeder, 4/0 copper 3-phase lines with no. 6 copper single-phase laterals

veloped by the Northern States Power Company.

The data obtained in these tests indicated, however, that a heavy-duty auxiliary protector, which would be adequate for this type of joint-use construction, could be provided. Fig. 7 shows an experimental model which was tested and showed possibilities of meeting minimum performance standards.

Time-current curves derived from test data for the standard *99A* protector and the experimental model are shown in Fig. 8. These curves are based on initial values of fault current and only approximate a constant-current condition. They show currents that can be carried for indicated times without incapacitating the protector.

The time-current curve for *109H* steel wire is also shown in Fig. 8. This curve is based upon initial current values.

An important effect which was noted was the increase in resistance of the telephone-line wire due to heating. This was especially noticeable on the lower values of fault current where the fusing occurred after a relatively long period of time or did not occur at all. This phenomenon is discussed in more detail in the following.

Application of Test Results

The results of the tests had two important applications. The first is in the design of a new protector which will withstand the duty cycle imposed upon it by high-capacity 12.5-kv systems. The performance of the Bell Laboratories' experimental model was particularly encouraging in this respect. It is anticipated that the new protector will be better under most fault conditions than the types of wire normally used for exchange open-wire plant so that a path to ground will be provided until the telephone wire fuses.

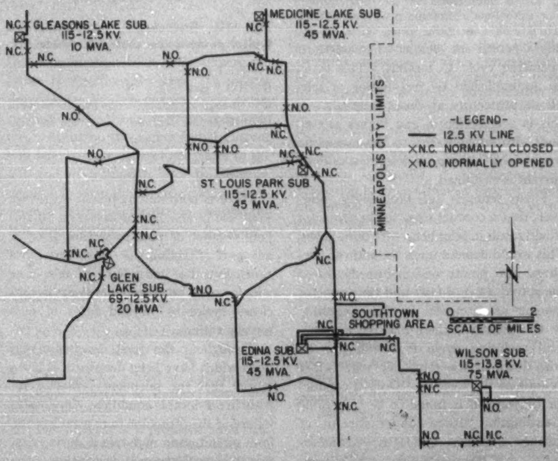
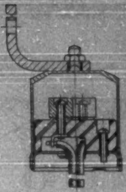


Fig. 3. Typical suburban area showing 12.5-kv main-line feeders

Fig. 4. Heavy-duty auxiliary telephone protector, type 99A



The other principal application of the test results is in the design of protection schemes for 12.5-kv distribution feeders. A simple example of a distribution feeder is shown in Fig. 9 and the curves associated with the various power protective devices are given in Fig. 10. Fig. 11 shows the curves of the power and telephone devices superimposed and the areas where co-ordination can and cannot be obtained may be seen. (These curves are for constant current conditions and do not take into consideration any current decay.)

For example, protection co-ordination is usually not possible in the main line of section *A* of Fig. 9. Since the station breaker will not respond to minimum fault currents (20-ohm faults), only burn-off of the telephone wire in the fault circuit in the event of low-current faults or operation of the station breaker on sustained high currents are available to open the circuit. Burn-off of a telephone wire can occur only when the cold resistance of the telephone wire is small, and where its increasing resistance does not combine with the source impedance to prevent fusing before the recloser can complete its cycle. This condition can prevail only over a relatively short distance from a telephone protector. Even if burn-off is acceptable as a means of de-energizing the contact, the proportion of section *A* in which this can occur is so small that joint use cannot, as stated, be permitted anywhere in this section. Open-wire joint use is possible on the fused laterals in this section, however, since the fuses will clear any faults on the laterals practically instantaneously. Likewise, open-wire joint use is possible on the taps protected by the 70-ampere reclosers. In this case, the minimum fault current would have to be checked against the recloser time-current cycle.

In section *B*, the minimum fault current problem on the main-line portion (protected by the 140-ampere recloser) is different from the main line in section *A* in that a reduction in the 20-ohm fault resistance to some lower value, with its corresponding increased fault current,

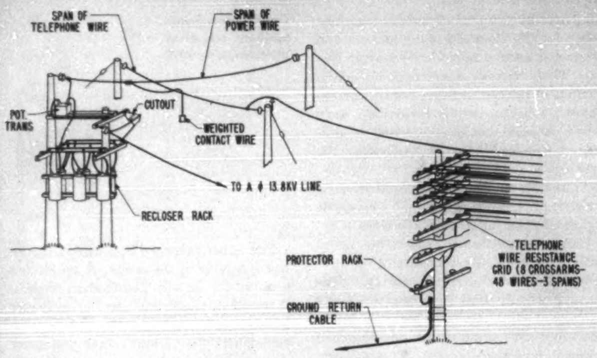

SPAN OF TELEPHONE WIRE
SPAN OF POWER WIRE
POT. TRANS
CUTOUT
WEIGHTED CONTACT WIRE
TO A ϕ 13.8KV LINE
RECLOSER RACK
PROTECTOR RACK
TELEPHONE WIRE RESISTANCE GRID (8 CROSSARMS-48 WIRES-3 SPANS)
GROUND RETURN CABLE

Fig. 5. Test setup

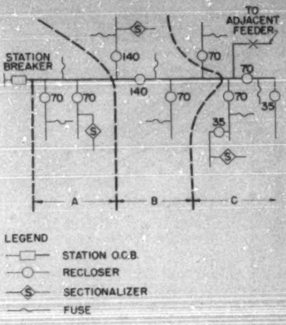
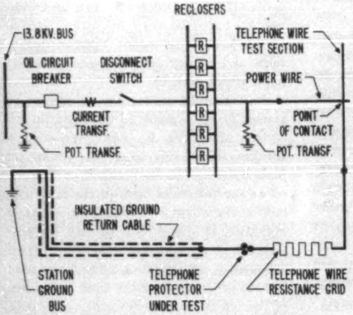

TO ADJACENT FEEDER
STATION BREAKER
140
140
70
70
70
70
70
70
35
35
A
B
C
LEGEND
STATION O.C.B.
RECLOSER
SECTIONALIZER
FUSE

Fig. 9. Simplified version of 12.5-kv feeder


RECLOSERS
13.8KV. BUS
OIL CIRCUIT BREAKER
DISCONNECT SWITCH
CURRENT TRANSF.
POT. TRANSF.
TELEPHONE WIRE TEST SECTION
POWER WIRE
POINT OF CONTACT
POT. TRANSF.
INSULATED GROUND RETURN CABLE
STATION GROUND BUS
TELEPHONE PROTECTOR UNDER TEST
TELEPHONE WIRE RESISTANCE GRID

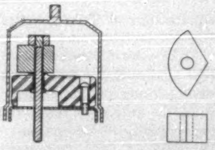


Fig. 6 (left). Diagram of test circuit

Fig. 7 (above). Experimental-model heavy-duty auxiliary telephone protector developed by Bell Telephone Laboratories

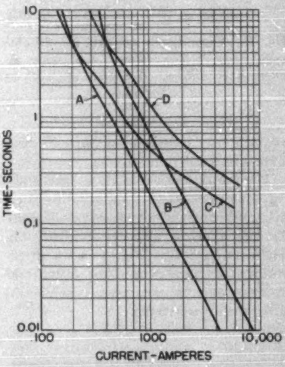


Fig. 10. Time-current characteristics of typical power-system protective devices

A—50 T fuse
B—100 T fuse
C—70-ampere recloser: total time to lock-out
D—140-ampere recloser: total time to lock-out

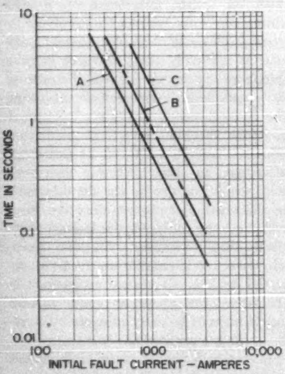


Fig. 8. Time-current characteristics of telephone wire and protective devices

A—99A protector
B—109H steel line wire: fusing characteristics
C—Experimental model protector

might permit the recloser to complete its operating cycle to lockout. This could be accomplished by installation of telephone protectors at closer intervals. If this is accomplished, any higher ground fault current value up to the maximum would lock out the recloser if the fault current is sustained.

There remains the probability, however, that a contact close to the protector would result in telephone-wire burn-down. This would depend upon the ratio of telephone and power system impedances at the actual point of fault and the protector spacing as discussed in the preceding. In this situation there may be no practical reclosing cycle that will prevent burn-down of the telephone wire before lockout of the recloser. Only if conductor burn-down is found to be acceptable co-ordination (from the standpoint of economics, safety, and other considerations), can open-wire joint use be considered in this situation.

There is nothing to prevent open-wire joint use in any lateral beyond a 70-ampere recloser or fuse in section *B* or in any part of section *C*. Tests were quite conclusive that burn-down of telephone wire is not to be expected in the faster recloser cycles or beyond fuses of customary ratings.

Throughout the fault clearing operations which have been described, it is essential that the telephone protector remain in serviceable condition. Obviously, failure of the protector to re-establish the fault circuit or to maintain it during current flow would prevent completion of the recloser cycle to lock-out.

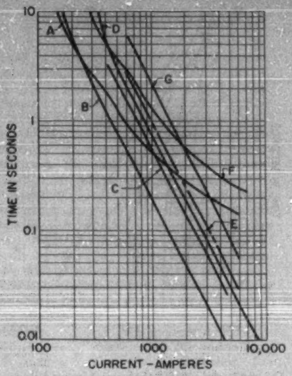


Fig. 11. Time–current curves showing co-ordination of power and telephone protective devices

A—70-ampere recloser: total time to lock-out
B—50 T fuse
C—99A protector
D—100 T fuse
E—109H steel line wire: fusing characteristics
F—140-ampere recloser: total time to lock-out
G—Experimental model protector

It should be pointed out that the curves of Fig. 11 should be used for the layout of co-ordinated protection schemes only if the phenomenon of current decay is recognized and taken into consideration. Fig. 12 shows a typical current-decay path plotted from test data for *109H* steel wire with an initial fault current of 410 amperes. It can be seen that the co-ordination with a 70-ampere recloser is questionable since the current drops below the time–current curve of the recloser before the recloser reaches lock-out. Actually, the higher initial current will tend to shift the recloser curve to the left and co-ordination may be possible, depending upon the specific circuit conditions. Further information concerning this characteristic is given in the appendix.

Conclusions

The following conclusions are based on the results of these tests:

1. The need for a heavier duty auxiliary telephone protector is greater than originally anticipated.

2. Development of a protector with time-current characteristics better than the characteristics of the telephone-line wire normally used on joint-use lines is feasible.

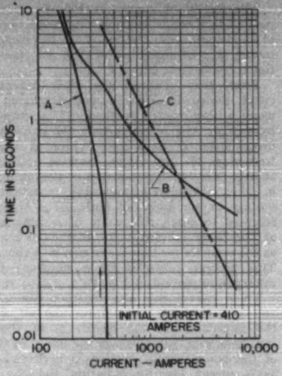


Fig. 12. Current-decay curve from actual test data showing questionable co-ordination with recloser

A—Path followed by decreasing current due to wire heating (109H wire)
B—70-ampere recloser: 13 AB total time to lock-out
C—Fusing current 109H wire (initial current)

3. Consideration should be given to the protection derived from the removal of the supply voltage from the telephone plant by burn-off or fusing of the telephone conductor when prompt de-energization of the power circuit is impracticable.

4. An important characteristic which had not previously been recognized was the fault-current decay resulting from the increase in resistance due to heating of the telephone-wire portion of the circuit. This characteristic is the controlling factor in determining fault-current magnitudes and durations in many situations and is increasingly important on high-capacity distribution systems. Future investigation of this phenomenon is necessary in order to evaluate its effect on co-ordination of power and telephone protective devices.

5. Minimum as well as maximum values of power-system fault currents and their duration as well as maximum load currents must be controlled if satisfactory co-ordination is to be achieved without affecting normal service standards of both power and telephone systems.

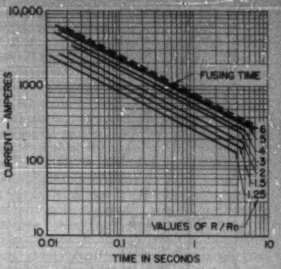


Fig. 13. Calculated time–current curves of 109H steel telephone wire for constant R/R₀ ratios

R_0 = resistance at 20 C
R = final resistance

6. High-capacity higher-voltage power-distribution systems generally involve a wide range of fault currents in addition to larger load currents. This, together with the more frequent use of automatic reclosing on faults, requires the closest co-operation between power and telephone engineers in the design of joint-use plant. Any compromise which is necessary to permit open-wire joint use must not be at the expense of co-ordinated protection.

Appendix. Calculated Curves of Current Decay in Steel Wire

It is desirable in many cases to apply the largest circuit recloser with delayed settings that a power-telephone contact will permit. To accomplish this, it is necessary to determine as accurately as possible the minimum fault current which the fault circuit can sustain throughout the reclosing cycle to lock-out. In order to establish such lower limits for one type of commonly used telephone-line wire, the relative change in resistance of *109H* steel line wire was calculated as a function of time for a number of fixed current values. The following energy equation was used:

$$I^2R_0[1+\rho(\Delta T)]t=\frac{WC_M(\Delta T)}{0.239}$$

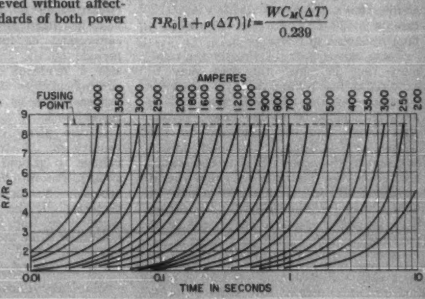


Fig. 14. Resistance ratio versus time of 109H steel telephone wire for constant currents

R_0 = resistance at 20 C
R = final resistance

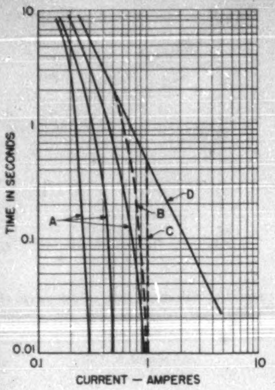


Fig. 15. Calculated constant-current fusing curve and maximum current-decay curve for 109H steel telephone wire

A—Maximum-current decay (constant voltage at fault)
B—Intermediate current decay
C—Sustained current (zero fault resistance)
D—Fusing time for sustained currents

where

I = assumed sustained current value in amperes
R_0 = resistance of unit length of wire at 20 degrees centigrade (C) in ohms
ρ = thermal coefficient of resistivity of material
ΔT = increase in temperature during interval
t = time interval in seconds
W = weight of unit length of wire in grams
C_M = specific heat of material

The equation was solved for t for various values of I assuming that ρ and C_M are constants for each temperature interval.

Energy losses from radiation and convection were neglected, so the resulting curves show minimum limits rather than actual values. The effects of the galvanizing were also neglected. Fig. 13 shows the results of these calculations.

It should be noted that the calculations assume that the wire will fuse at a temperature of 1,550 C. This corresponds to a final resistance of 8.5 times R_0.

These curves indicate directly the minimum time in which this wire may be expected to fuse for any fault for which the impedance of the telephone wire itself is negligible compared to the power system impedance. This can occur only in the section of telephone circuit immediately adjacent to the protector, generally within 200 feet. In this interval, the circuit recloser must complete its cycle to lock-out in less time than indicated by the curves if fusing of the telephone-line wire is to be prevented.

The family of curves in Fig. 14 were plotted as a convenience in the calculation of the current-decay curves. They represent merely a rearrangement of the information presented in Fig. 13.

From these curves, the time–current curve shown in Fig. 15 was developed, using a step-by-step method. This step-by-step method assumes that the current drops instantly to a lower value at the beginning of any time interval. The change in resistance to arrive at this value is then calculated assuming constant voltage at the point of fault. The time required to reach this value of resistance at the lower current value is then taken as the ordinate for that lower current value. This process, repeated for a number of intervals, is plotted in Fig. 15.

These curves are also limiting values, and arise from the assumption that all of the impedance in the circuit is in the telephone wire. In this case the impedance of the power circuit portion is neglected, as is energy loss by radiation and convection. These curves indicate the maximum current decay which can take place because of heating and the resultant increase in resistance of the telephone wire. Actual current values as a function of time will vary from sustained initial currents to the maximum decay shown, depending upon the relative magnitude of the fixed power-circuit impedance and the increasing telephone-wire resistance.

The curve for an intermediate rate of current decay is also shown in Fig. 15.

Example

For a fault midway between two protectors spaced at half-mile intervals on *109H* steel wire, the actual fault resistance in the telephone circuit is approximately 10 ohms at 20 C. With the assumption that only one protector operates, the initial current should be between 500 and 700 amperes (see Fig. 1), depending upon the impedance of the power system. At these values of initial current, the curves in Fig. 16 indicate that more than 200 amperes will be available to operate the circuit reclosers at the end of 5 seconds. It should be noted that burn-down or fusing of the telephone wire could not necessarily be prevented, depending upon the choice of recloser cycle for this particular application.

The use of these calculated curves tends to give pessimistic results with regard to de-energization of the power circuit since no allowance is made for heat loss from the wire by radiation and convection, or for cooling during open periods of the recloser cycle. There is also a tendency for the recloser time–current curve to shift to the left since the recloser curve is plotted on a constant-current basis.

If the power contact occurs in the vicinity of the telephone protector, the voltage to ground at the contact would always remain below the 3,000 volts necessary to break down an adjacent protector. If the fault occurs at a point more remote from a protector, the current-resistance drop along the telephone wire between the fault and the operating protector may rise to 3,000 volts or more, which may operate the adjacent protector. This would provide a parallel path to ground and would result in added current in the power-circuit protective device, which would provide an additional safety factor in operating the power system recloser, circuit breaker, or fuse.

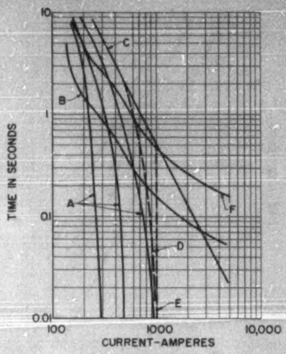


Fig. 16. Calculated fusing and current-decay curves for 109H steel wire with 70-ampere recloser curves

A—Maximum current decay (constant voltage at fault)
B—Intermediate current decay
C—Sustained current (zero fault resistance)
D—Fusing time for sustained currents
E—B curve for 70-ampere recloser
F—70-ampere recloser 13AB (total time to lock-out)

References

1. National Electrical Safety Code—Part 2. Handbook H32, National Bureau of Standards, Washington, D. C., 5th edition, Sept. 23, 1941.

2. Protection Features for the Joint Use of Wood Poles Carrying Communication Circuits and Power-Distribution Circuits above 5,000 Volts, J. O'R. Coleman, A. H. Schirmer. *AIEE Transactions (Electrical Engineering)*, vol. 57, March 1938, pp. 131–40.

3. Joint Use of Pole Lines for Rural Power and Telephone Services, J. W. Campbell, L. W. Hill, L. M. Moore, H. J. Scholz. *Ibid.*, vol. 66, 1947, pp. 519–24.

4. Some Aspects of Joint Use of Wood Pole Lines with Increasing Line Voltage, W. E. Bloecker, W. R. Bullard. *Ibid.*, vol. 72, pt. I, Nov. 1953, pp. 709–12.

5. Electrical Protection of Telephone Systems, W. R. Bullard, J. B. Hays, H. O. Saunders. *Ibid.*, vol. 73, pt. I, Sept. 1954, pp. 385–88.

6. Positive Disconnection of Distribution Circuits During Faults to Ground, Engineering Report No. 47. Joint Subcommittee on Development and Research, Edison Electric Institute, New York, N. Y., and the Bell Telephone System, New York, N. Y., July 24, 1941.

Discussion

Frank W. McStay (Line Material Company, Milwaukee, Wis.): The participants in the Minneapolis tests deserve to be complimented upon their valuable contribution.

At one point the paper comments on the burden imposed on the telephone protection system when automatic oil circuit reclosers are used, the burden imposed being the duration of fault current flow. It might be desirable to note that recloser operating characteristics today are quite versatile and that the burden imposed when reclosers are used would to a considerable extent depend upon the operating sequence and characteristic selected when applying the reclosers.

Whether ground-fault relaying should be avoided is perhaps a moot point. There are advantages to be gained by its use. As an example, a 100-ampere recloser will trip at 200 amperes. Assume that the normal load on such a recloser is 80 amperes. For ground trip pickup at 100 amperes the unbalance required for tripping would therefore be 125% of the regular load current. In other words, the minimum tripping current with ground-fault relaying would be only one-half that required when ground relaying is not employed. Yet the unbalance required to trip this particular recloser would be greater than the normal full-load current.

As a matter of terminology in connection with the fault current decay, whether the recloser time-current characteristic should be said to shift to the right or to the left depends on the reference fault current values held in mind, initial fault current, or that which exists sometime after the fault initiation. Shifting the characteristic to the right would indicate slower operation. And, of course, if the recloser characteristic as shown for constant currents is not shifted, then, because of the current decay involved in this application, the point on the characteristic which would indicate the actual recloser opening time would lie to the left of the initial fault current.

The time-current characteristic for the 140-ampere recloser shown in Fig. 11 should not extend past 6,000 amperes.

E. G. Albrecht, E. W. Christoferson, A. E. Dietz, and **J. C. Slothower:** Mr. McStay is quite correct in stating that a very wide choice of recloser operating characteristics is available to the power system designer. However, the choice of these characteristics and operating cycles is limited in any particular case by the operating requirements of the particular power feeder. Obviously, an increase in the number of fast operations will reduce the burden imposed upon the telephone protection system, but it introduces other problems in the power protection scheme.

With reference to the suggestion that ground-fault relaying be incorporated as a possible solution, it would appear that in the example chosen a 100-ampere recloser with an 80-ampere load, there would be no particular difficulties in protection co-ordination in this case. An 80-ampere load would involve maximum fault current of probably not much more than 1,200 amperes. With reference to Figs. 1 and 15, it can be seen that the maximum fault currents of this magnitude would, in general, involve neither burn-down nor an appreciable amount of current decay. Nevertheless, ground-fault relaying offers attractive possibilities and may permit protection co-ordination in many cases where it would not otherwise be possible.

The statement that the recloser curve shifts to the left is based upon the final value of fault current. If the initial value of fault current is considered, then the recloser will require a longer time to operate with current decay, and it could be considered that the operating curve shifts to the right. The authors feel that information concerning the amount of this shift would be extremely useful in laying out power system protection schemes involving open-wire telephone contacts. Allowances for this factor may permit the use of a larger recloser or longer operating cycle.

Co-ordinated Protection for Open-Wire Joint Use—Ontario Tests

H. M. ELLIS
ASSOCIATE MEMBER AIEE

J. W. PHELPS
NONMEMBER AIEE

C. L. ROACH
MEMBER AIEE

R. E. TREEN
ASSOCIATE MEMBER AIEE

WHEN POLES are jointly used by power and communication utilities, or the two types of plant are on separate pole lines but are in situations of close proximity, the possibility of accidental contact between the two types of plant must be recognized. The first line of defense against such occurrences is to minimize the probability of contact by means of adequate construction and clearance standards, as covered by the National Electrical Safety Code in the United States or the Canadian Electrical Code in Canada.

Notwithstanding the grade of construction employed, however, the possibility of accidental contact will still exist due to abnormal weather conditions or accidents. To ensure the safety of persons and to limit damage to plant and property in the event of contact, certain co-ordinated protection features are incorporated in the design of the jointly used plant. Positive and prompt interruption of the contact current to the communication plant, either by de-energization of the power circuit or by some other means, and limitation of the voltages on the communication plant to the range of safe operating characteristics of communication protective equipment, constitute the modern view of the objectives of co-ordinated protection.

When the telephone plant consists of open-wire conductors, and the power conductors to which the telephone plant is exposed are operated at a voltage-to-ground below 3 kv, the protection furnished by the conventional station protection apparatus is usually sufficient. The fuse at the subscriber's station protector is in series with a small air gap which breaks down to ground in the event of contact. The consequent flow of fault current through the fuse causes it to operate and isolate the station. However, in joint use, or occasionally at crossings, when the power supply conductors operate at greater than 3 kv, auxiliary protectors are connected between the communication wires and an effective ground at intervals along the line. If a multigrounded neutral is present on the line, the preferred ground connection for the auxiliary protectors is to the multigrounded neutral as illustrated in Fig. 1.

The gaps between electrodes in the model of these auxiliary protectors used heretofore (*99A*) are such that they break down at about 3,000 volts rms. This value was chosen in the design of the *99A* protector because voltages below this

---

Paper 56-238, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30-February 3, 1956. Manuscript submitted October 20, 1955; made available for printing December 7, 1955.

H. M. ELLIS and R. E. TREEN are with the Hydro-Electric Power Commission of Ontario, Toronto, Ont., Canada; J. W. PHELPS is with Bell Telephone Laboratories, Inc., Murray Hill, N. J.; and C. L. ROACH is with the Bell Telephone Company of Canada, Montreal, Que., Canada.

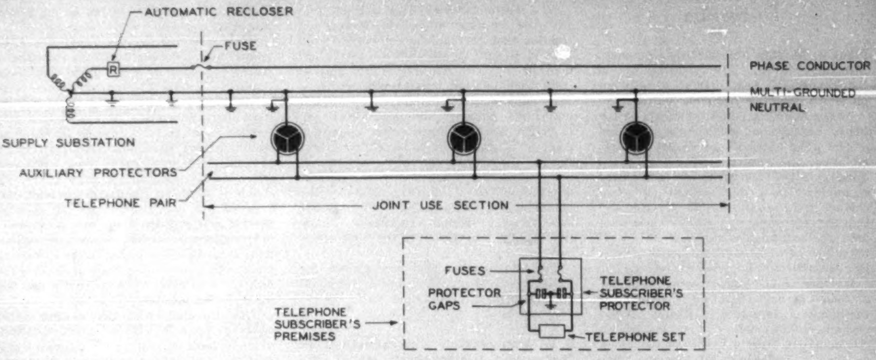

AUTOMATIC RECLOSER
FUSE
R
PHASE CONDUCTOR
MULTI-GROUNDED NEUTRAL
SUPPLY SUBSTATION
AUXILIARY PROTECTORS
TELEPHONE PAIR
JOINT USE SECTION
FUSES
PROTECTOR GAPS
TELEPHONE SUBSCRIBER'S PROTECTOR
TELEPHONE SUBSCRIBER'S PREMISES
TELEPHONE SET

Fig. 1. Elementary joint use section illustrating co-ordinated protection

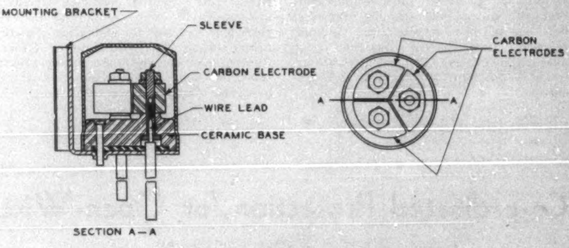

MOUNTING BRACKET
SLEEVE
CARBON ELECTRODE
WIRE LEAD
CERAMIC BASE
SECTION A—A
CARBON ELECTRODES
A
A

Fig. 2. Section and top views of 118A protector assembly

value will not maintain an arc in the subscriber's protector fuses after they have blown. Breakdown of the auxiliary protector follows the operation of the subscriber's protector, and effectively quenches the arc in the fuses.

A contact between a supply wire energized at over 3,000 volts-to-ground and a communication conductor results, therefore, in the breakdown of a near-by auxiliary protector gap, with consequent flow of power-system ground fault current. If the design of the co-ordinated protection scheme is sound, the impedance in the fault path, including the part of the communication plant involved, is low enough to permit the fault current to operate the power system protective device, such as a fuse or recloser, and hence effect prompt and positive de-energization. The low impedance in the communication part of the fault path, assured by the proper spacing of the auxiliary protectors, also limits the voltage appearing on the communication plant.

Need for Improved Protector

The auxiliary protector, like all such devices, has an inverse time–current characteristic beyond which its destruction takes place. Obviously, under the present co-ordination scheme, this characteristic must exceed that of the power-system protective device or the protector will be destroyed before the power device operates, and de-energization will not be realized.

With the trend towards greater loads on power distribution circuits, the ratings of fuses and other protective devices have, in many cases, risen to the point where co-ordinated protection is unattainable with the telephone auxiliary protector used heretofore. This is illustrated by the fact

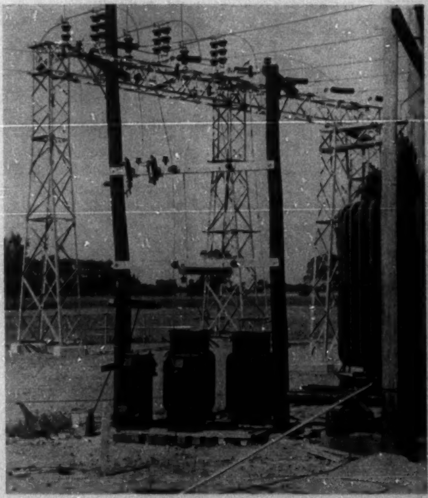


Fig. 3. Transformers and disconnect switches used in tests

that line-to-ground fault currents for modern rural distribution stations, such as those of The Hydro-Electric Power Commission of Ontario, will vary from 1,000 to 9,000 amperes, depending upon the characteristics of the supply system. The power protective devices normally used on these rural distribution systems have time-current characteristics which often exceed the current-carrying capabilities of the *99A* auxiliary protectors.

Rural distribution lines in Ontario, which are fed from modern distribution stations varying in capacity from 2,000 to 4,000 kva at 8 to 12.5 kv, have coordinated protection based upon a combination of reclosers, sectionalizers, and fuses. The trend is to use reclosers on feeders and sectionalizers on laterals, especially for long lines where fuse maintenance is costly. Generally, reclosers, rated from 50 to 100 amperes, and set for one fast and three retarded operations, are used on rural feeders.

Several years ago it became apparent that the savings and other benefits of the joint use of poles could no longer be fully realized unless a better telephone auxiliary protector could be provided.[1] The conflict between increasing loads on the power system and the economic desirability of jointly used plant finally progressed to the point where Ontario Hydro realized that an improved telephone protector was required in the near future or jointly used plant would have to be curtailed and existing joint use reduced. Accordingly, in 1952, Ontario Hydro suggested to The Bell Telephone Company of Canada and Bell Telephone Laboratories that the possibility of supplying an improved protector be investigated. On the basis of similar suggestions arising in the United States, a series of field tests was made in Minneapolis, Minn., in 1953 to evaluate certain aspects of the problem.[2]

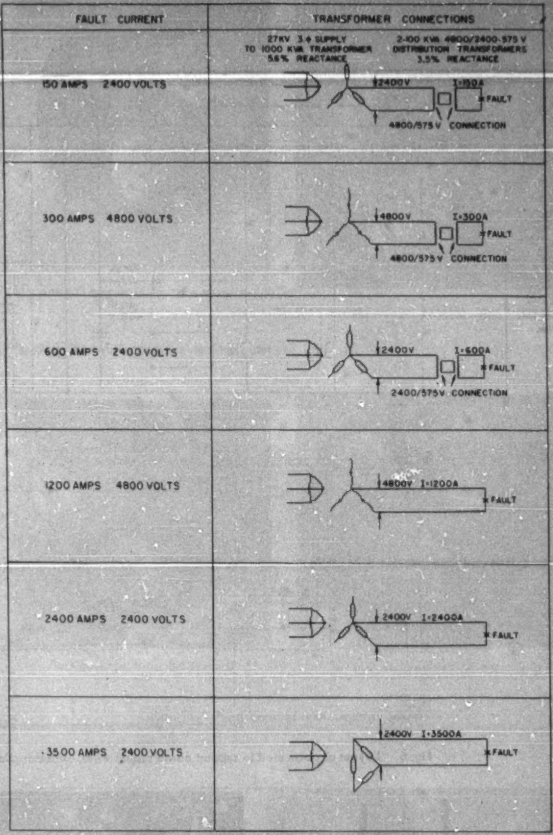

FAULT CURRENT
TRANSFORMER CONNECTIONS
27KV 3 ϕ SUPPLY TO 1000 KVA TRANSFORMER 5.6% REACTANCE
2-100 KVA 4800/2400-575 V DISTRIBUTION TRANSFORMERS 3.5% REACTANCE
150 AMPS 2400 VOLTS
2400V
I=150A
FAULT
4800/575 V CONNECTION
300 AMPS 4800 VOLTS
4800V
I=300A
FAULT
4800/575 V CONNECTION
600 AMPS 2400 VOLTS
2400V
I=600A
FAULT
2400/575V CONNECTION
1200 AMPS 4800 VOLTS
4800V I=1200A
FAULT
2400 AMPS 2400 VOLTS
2400V I=2400A
FAULT
3500 AMPS 2400 VOLTS
2400V I=3500A
FAULT

Fig. 4. Power transformer connections used in tests

New Protector Design

Part of the Minneapolis tests were devoted to the determination of the time-current characteristic of the standard *99A* protector for times and currents in line with the modern demands of joint use. The results indicated that the *99A* was inadequate for many of the applications desired. Preliminary models of a new protector were also tested at that time and as a result, the conclusion was reached, that a protector, adequate for present-day conditions, could be developed.

Realizing the necessity of providing a protector capable of handling the necessary currents long enough to perform its intended function, the engineers at the Bell Telephone Laboratories therefore produced a new design along the lines indicated by the tests at Minneapolis.

This new protector in its final form has been coded the *118A* protector and will replace the *99A* protector in the Bell System. It consists essentially of three wedge-shaped carbon electrodes arranged on a ceramic base as illustrated in Fig. 2. The gaps in the new protector are adjusted so that breakdown occurs at approximately 2,000 volts rms. For normal operation, an electrode is connected to each of the conductors of a telephone open-wire pair, and the third electrode to an effective ground, such as a power system multigrounded neutral (see Fig. 1).

Protector Test Layout

In order to determine the time-current characteristic of the new protector, a series of tests was performed cooperatively by Bell Telephone Laboratories, The Bell Telephone Company of Canada, and The Hydro-Electric Power Commission of Ontario at the Ontario Hydro transformer station near Brampton, Ont., during the summer of 1954.

This station, known as the Pleasant Transformer Station, was a new installation with only one feeder in service, and space was available in the relay building for all controls and equipment required

(*Continued on page 230*)

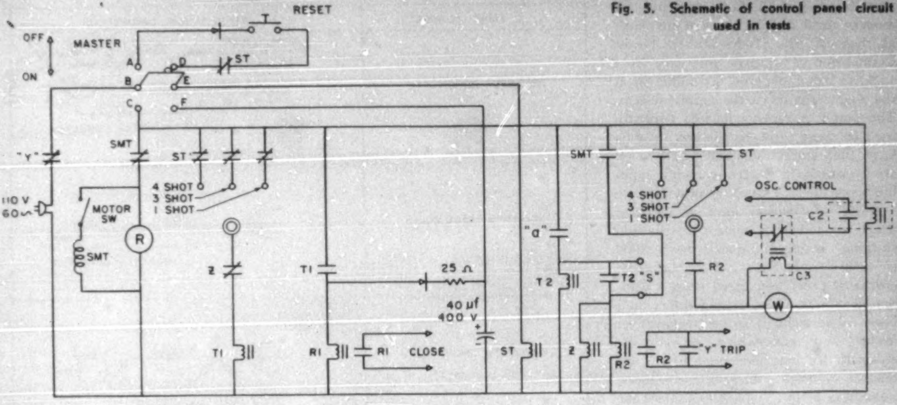

RESET
OFF
ON
MASTER
A
B
C
D
E
F
ST
"Y"
SMT
110 V
60~
MOTOR SW
SMT
R
4 SHOT
3 SHOT
1 SHOT
Z
T1
T1
25 Ω
40 μf
400 V
R1
R1
CLOSE
ST
"a"
T2
T2 "S"
Z
R2
R2
"Y" TRIP
OSC. CONTROL
C2
C3
W

Fig. 5. Schematic of control panel circuit used in tests

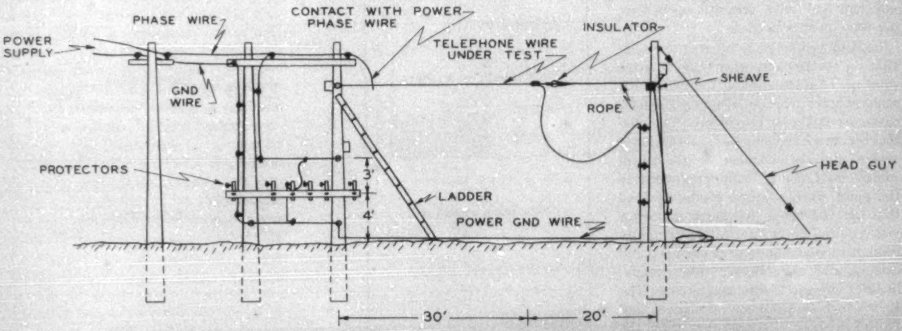

POWER SUPPLY
PHASE WIRE
GND WIRE
CONTACT WITH POWER PHASE WIRE
TELEPHONE WIRE UNDER TEST
INSULATOR
SHEAVE
ROPE
PROTECTORS
3'
4'
LADDER
POWER GND WIRE
HEAD GUY
30'
20'

Fig. 6. Layout of poles used to support power supply wires, protector samples, and wire test samples

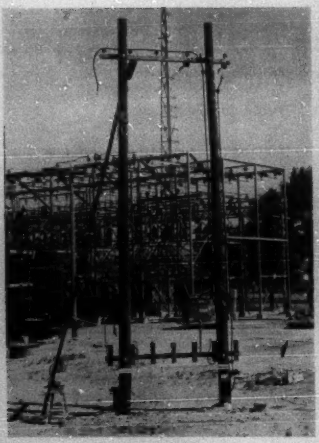
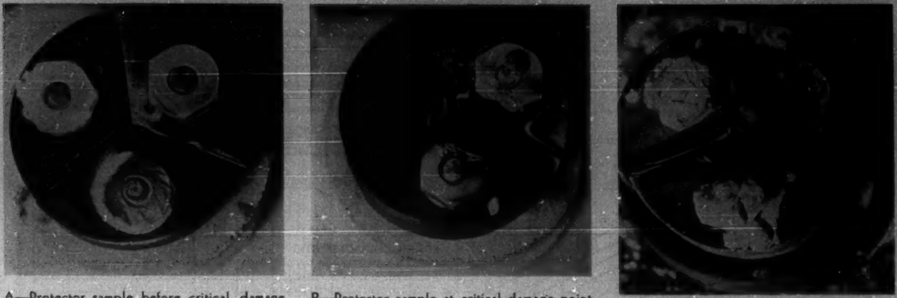


Fig. 7 (left). Two-pole H-structure used to support test specimens and power supply wires

Fig. 8 (above). Test on one of the protector samples



A—Protector sample before critical damage point

B—Protector sample at critical damage point

C—Protector sample beyond critical damage point

Fig. 9. Illustrations of critical damage point

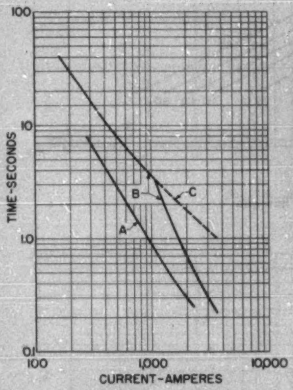


Fig. 10. Comparison of time–current characteristics of two telephone protectors having different types of electrode material

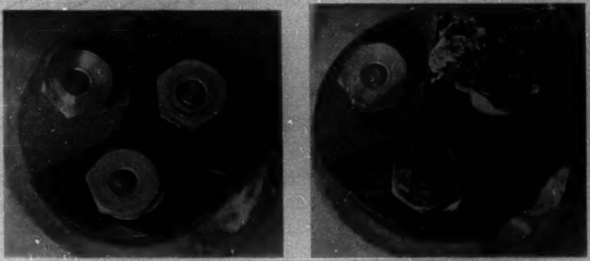
A—118A protector with low-graphite-content electrodes

B—118A protector with no. 10 American Wire Gauge copper leads and high-graphite-content electrodes

C—118A protector with 0.165-inch copper leads and high-graphite-content electrodes

Table I. Characteristics of Line Wires Tested

Kind of Wire	Nominal Diameter, Inch	Breaking Strength, Pounds-Min.	Weight per Mile, Pounds	Electrical Resistance When New. Ohms per Loop Mile, Approximate
H steel	0.109	1,213	170	70
Copper-steel	0.104	1,325	158.5	25
Copper	0.104	550	173	10



A—600 amperes for 6.6 seconds

B—600 amperes for 6.95 seconds

Fig. 11. Illustrations of arc transition point

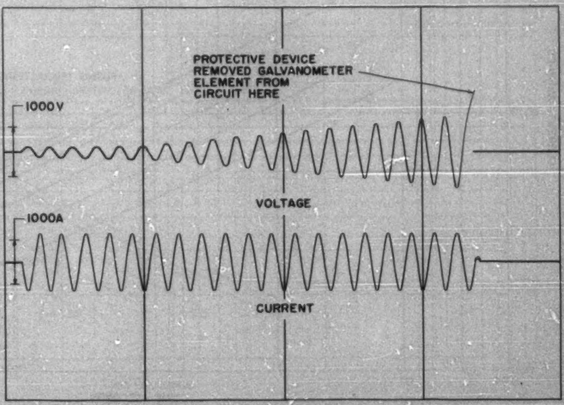


Fig. 12. Voltage and current traces of 109H steel line wire at 1,300 amperes

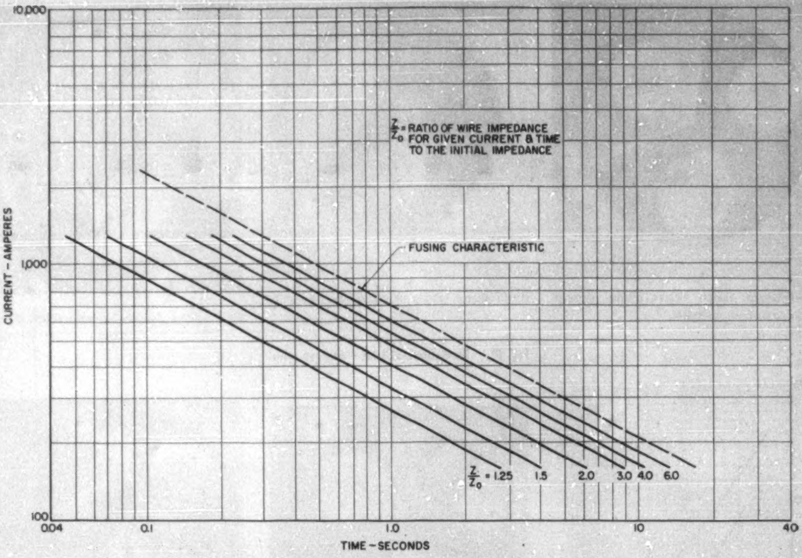


Fig. 13. Impedance and fusing characteristics of 109H steel line wire

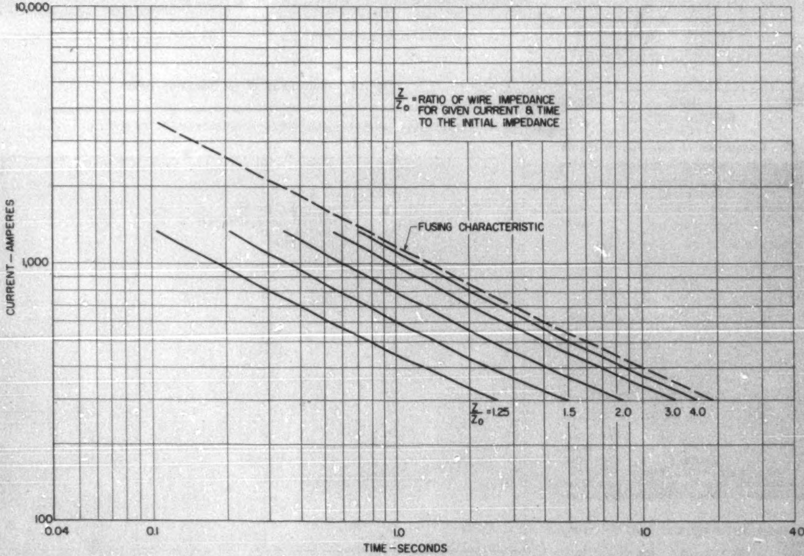


Fig. 14. Impedance and fusing characteristics of 104 copper-steel line wire

Vol.
Pt. 1
1956

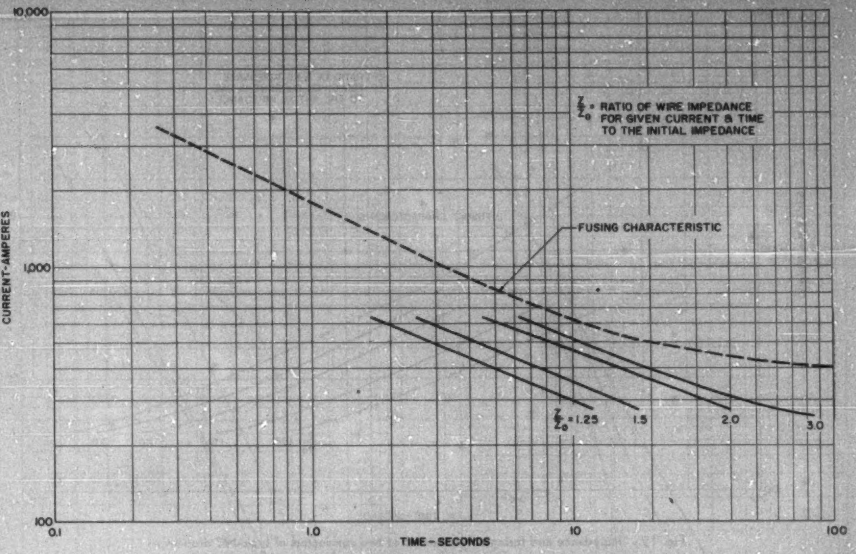


Fig. 15. Impedance and fusing characteristics of 104 copper line wire

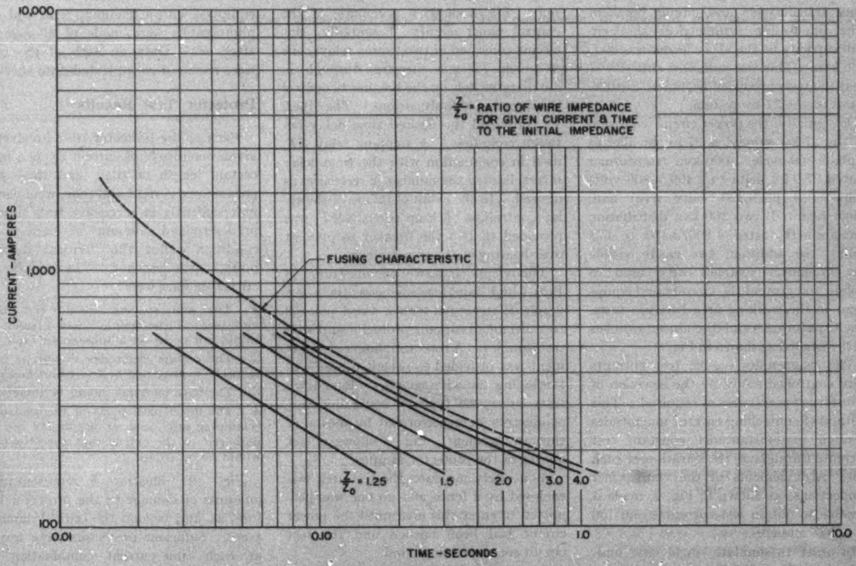


Fig. 16. Impedance and fusing characteristics of one conductor of type-NC drop wire

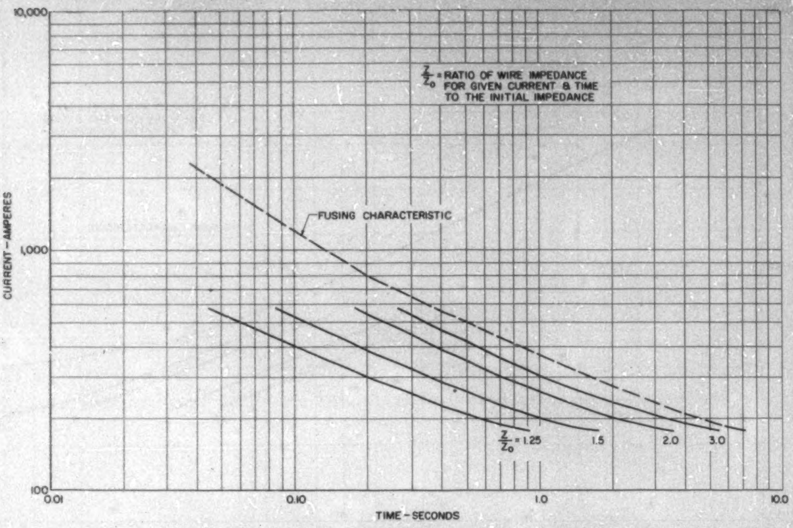


Fig. 17. Impedance and fusing characteristics of two conductors of type-NC drop wire

(*Continued from page 225*)

for the tests excluding, of course, the outdoor transformers and circuit breakers. The line-to-line symmetrical short-circuit capacity on the 27-kv feeder was 300,000 kva. This was sufficient capacity so that the tests produced negligible disturbances on the 27-kv system.

Essentially, the power circuit consisted of the 27-kv supply, a 27-kv breaker, a 3-phase 60-cycle 1,000-kva transformer (rated 27.6-kv delta to 2,400/4,800 volts delta or 4,160/8,320 volts wye) and, when required, two 100-kva distribution transformers (rated 4,800/2,400 to 575 volts). In addition, two easily visible air disconnect switches were used to isolate and ground the circuits and equipment in the working area between tests. The transformers and disconnect switches were installed as shown in Fig. 3.

The magnitudes of the test currents were controlled solely by the insertion of transformer reactances as required. This method of controlling current magnitudes resulted in substantially constant test currents throughout the duration of each test. Arrangements of the transformer connections, as shown in Fig. 4, made it possible to obtain test currents from 150 to 3,500 amperes.

In order to simulate single- and multiple-shot characteristics of distribution fuses and oil circuit reclosers, a special control panel was constructed for the 27.6-kv breaker used to switch the fault current. Fig. 5 shows a schematic of the control panel circuit. Essentially, this circuit consisted of two timing relays, one of which, *T2*, was energized through an auxiliary switch on the breaker to control the duration of fault current. The other, *T1*, provided the desired time delay between reclosures. A stepping relay *ST*, used in conjunction with the time-delay relays, limited the number of reclosures as desired. In the event of failure of the relay controls, a back-up relay, *SMT*, was provided to trip the breaker to prevent over-heating of the transformers.

The telephone network is illustrated in Figs. 6 and 7 and consisted essentially of a 2-pole H-structure with a third pole 50 feet distant to support protector and wire test samples. A rope and sheave arrangement was provided to permit raising and tensioning the wire samples. Burndown tests on several types of communication conductors were performed by means of this arrangement. Fig. 8 shows a test on one of the protector samples.

As a safety measure, the test area was enclosed by a fence and no one was permitted to enter this area until the power circuit had been opened and the test circuit connected to ground.

Current and voltage measurements were made on two oscillographs. A photographic record of each test was made on a 16-millimeter movie camera equipped with a telephoto lens, and still photographs were made of all samples after test. Data on each of the tests were recorded on a printed data sheet.

Protector Test Results

Each of the protector tests involved a given magnitude of current applied for a certain length of time, and these two parameters, current and time, were varied independently in successive tests until a predetermined amount of damage, a condition called the "critical damage point," was reached. The criteria of critical damage were:

1. The gap spacing should not have increased to the point where breakdown would not occur on a subsequent fault.
2. The carbon electrodes should be held securely in place and be free from cracks.
3. The base material should be intact.
4. The mechanical parts of the protector (clamping nuts, and so on) should not be damaged to the extent that the structure would be weakened.

Fig. 9 illustrates representative amounts of damage to the protector before, at, and beyond the critical damage point. Sufficient protectors were tested at each time-current combination to provide adequate sampling.

(*Continued on page 232*)

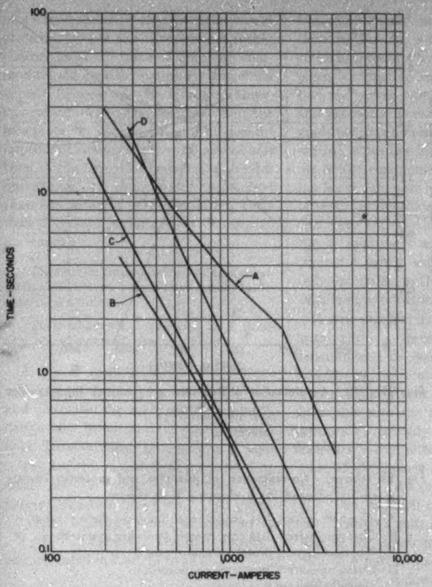


Fig. 18. Comparison of protector and wire characteristics

A—118A protector with no. 8 American Wire Gauge copper leads
B—99A protector
C—109H steel line wire
D—104 copper-steel line wire

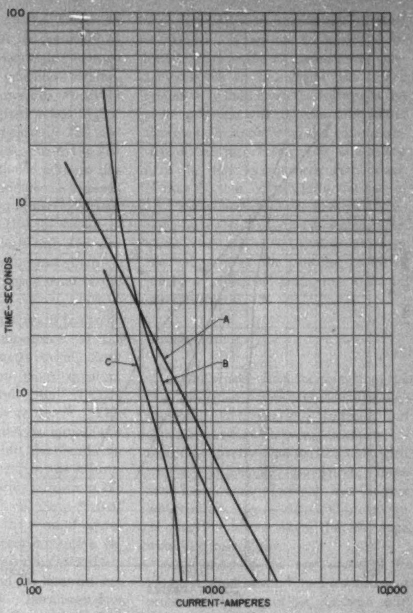


Fig. 19. Co-ordination of line-wire and fuse-link characteristics

A—109H steel-line-wire
B—100-ampere-fuse
C—Fault-current as a function of time

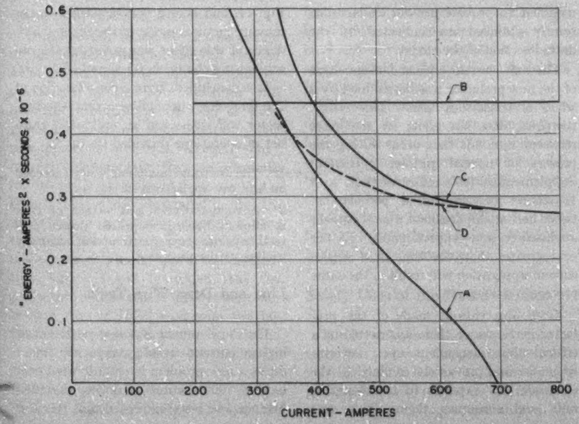
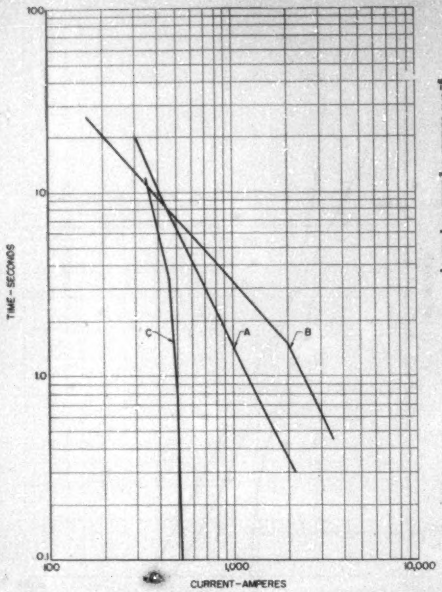


Fig. 20. Co-ordination of line-wire and fuse-link characteristics adjusted for a changing fault current

A—System energy input
B—109H steel-line-wire characteristic
C—Constant-current 100-ampere fuse characteristic
D—Changing-current 100-ampere fuse characteristic



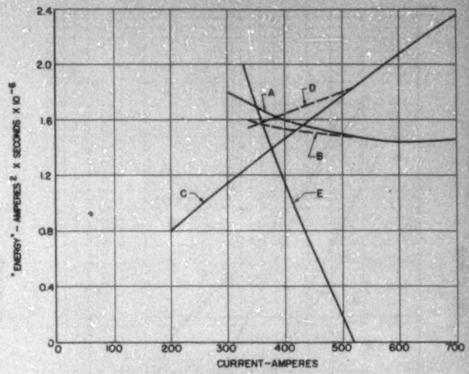


Fig. 21 (left). Co-ordination of line-wire and protector characteristics

A—104 copper-steel line-wire
B—118A protector
C—Fault current as a function of time

Fig. 22 (above). Co-ordination of line-wire and protector characteristics adjusted for a changing fault current

A—Constant-current 104 copper-steel line-wire characteristic
B—Changing-current 104 copper-steel line-wire characteristic
C—Constant-current protector characteristic
D—Changing-current protector characteristic
E—System energy input

(*Continued from page 230*)

The new protectors intended for these tests were manufactured with the same type of electrode material and the same size of lead-in wires as used in the standard *99A* protector. It was apparent, after very few tests, that the new protectors were not performing as well as the experimental models tested previously at Minneapolis. Upon investigation, it was discovered that the only significant difference between the two lots was in the type of carbon employed. The electrodes used in the construction of the experimental models had been made from carbon which was softer and contained a much higher percentage of graphite than does the standard carbon that is used in all other Western Electric Company protectors. In order to determine the precise difference in the two types of carbon, a quantity of electrodes was fabricated of the graphitized carbon and time–current characteristics were determined for both types. The results are plotted in Fig. 10, and indicate the marked superiority of the graphitized carbon as electrode material in high-current applications where a relatively large gap is involved and flaking of the electrode surface is therefore not a serious problem.

It will be noted that there is a discontinuity in the graphitized-carbon electrode curve at approximately 1,000 amperes. At and beyond this current point the lead-in wires fused before the electrodes or clamping nuts were damaged appreciably. In an attempt to determine the capability of the electrodes beyond the fusing point of the leads, some of the protectors were modified to provide heavier leads. The results obtained are indicated by the dotted section of the curve.

The time–current critical damage curve of the new protector was determined by a series of single-shot tests. This curve, therefore, does not allow for the short recovery intervals that occur during the process of normal recloser operation. Supplementary tests indicated that, with 1.5-second de-energization periods between current applications, a total current application time approximately 10 per cent greater than the time of a single-current application will result in the same protector damage.

From observations made of the protector performance there appears to be a critical time, dependent upon current, when the arc, previously confined in the electrode gap, expands to the clamping nuts and consumes them. This was arbitrarily called the "arc transition point" and occurs at approximately 90 per cent of the time required for critical damage. In other words, destruction of the protector is not a linear function of time but occurs rather suddenly after the protector has carried current satisfactorily for some period of time. This effect is illustrated on the protectors shown in Fig. 11, and is also very apparent in the moving pictures made of the tests. Evidence of this effect was noted by the observers as a change in the sound of the arc.

It is concluded, from a consideration of the foregoing phenomenon, that the protector will withstand an unlimited number of operations provided that:

1. The arc transition point is not exceeded on any one application of current.
2. A comparatively long period of time is allowed between repeated power protective-device sequences to permit adequate cooling of the protector.

Line and Drop Wire Tests

The development of a new protector of higher current-carrying capacity introduces a new problem in co-ordinated protection. Data obtained during the 1953 Minneapolis tests[2] indicated that the new protector capability characteristic would probably exceed the time–current fusing characteristics of some of the more commonly used communication exchange line wires. Because of the increased time–current characteristics anticipated in the new protector applications, the change-in-impedance characteristic with current and time of the line wire becomes of greater importance in co-ordinated protection than has hitherto been the case. Tests were therefore performed in Ontario to determine:

1. The simple time–current fusing characteristic of the line wires under steady fault-current conditions.
2. Their change-in-impedance as a function of current and time.

A typical oscillogram showing voltage and current traces is given in Fig. 12, and illustrates the basic test data used to derive the fusing and change-in-impedance characteristic.

Tests of this nature were made on drop wire and on steel, copper-steel, and copper telephone line wires. The physical characteristics of the line wires tested are given in Table I. The data obtained are presented in Figs. 13 to 17.

Engineering Applications

The advantages derived from the use of heavier leads, as shown by Fig. 10, resulted in the specification of no. 8 American Wire Gauge copper leads for the *118A* protector rather than no. 10 leads as are used in the *99A*. Several minor modifications of the protector design have been made since the field tests in an effort to improve the characteristic of the protector at currents below the fusing point of the no. 8 leads.

It will readily be seen by a comparison of the characteristics for the *99A* and the *118A* protectors (Fig. 18) that the development of the new protector will broaden considerably the field of joint power and communication facilities. The new protector provides, in fact, time–current characteristics that are adequate for the great majority of joint-use situations. This is also illustrated in Fig. 18, where the protector characteristic is contrasted to the fusing characteristics of the more commonly used types of communication wire. It will be noted that the protector characteristic everywhere lies beyond that of *109H* steel wire and beyond *104* copper-steel wire except in the range of currents below approximately 400 amperes.

Present practice, as embodied in the National Electrical Safety Code in the United States and the Canadian Electrical Code in Canada, is that, in the event of accidental contact between higher-voltage power conductors and telephone conductors, the power supply must be positively and promptly de-energized unless higher grade construction is employed. The separation of the two facilities due to the fusing of the communication conductor has not heretofore been considered as a means of securing co-ordinated protection under the codes. Consequently, the engineering of joint-use situations with the new protector would, under present practice, only extend up to the fusing point of the telephone conductor.

A new problem in this regard has to do with the change-in-impedance of the communications conductor during fault conditions; a problem more apparent owing to the longer fault duration times allowable with the new protector. This means that not only must the initial fault currents be computed, but also the magnitude of the change-in-current as a function of time. An example is given in the appendix in which this changing current effect causes a shift in the point of operation of the power system protective device to a part of its operating characteristic where burndown of the line wire appears to be a possibility. Burndown does not actually occur, but the example illustrates the necessity of a check of each joint-use section.

The point has been raised in discussions of joint-use engineering problems that in certain well-examined situations burndown of the communications conductor might be accomplished with the assurance of the same degree of safety as de-energization of the power supply would afford. As yet, no field studies have been instituted to prove or disprove this and, in addition, the economics of the problem have not as yet been examined. Such a scheme would again increase the field of joint-use endeavor.

The second example given in the appendix shows an engineering computation based on the assumption that burndown of the communications conductor is found to be a satisfactory alternative to de-energization. In such a case assurance is needed that the line wire will, in fact, fuse before a predetermined amount of damage is done to the protector.

Safety Uses of the New Protector

The breakdown voltage of the new protector has been set at approximately 2,000 volts to permit its use on lower voltage power systems to reduce the likelihood of power contact voltages remaining on the open wire and presenting a possible hazard to linemen. There are no special problems in engineering its use on lower voltage systems to de-energize the power system or to hold down the voltage on the communication wire to a reasonable value. However, the application and administration of this protector purely as a safety item has not as yet been established. It may be that certain lines, because of age, tree conditions, and weather conditions, have a bad history of power contacts. These, then, would be the cases where, in the judgment of the joint-use engineers involved, additional protection would be of some benefit.

Conclusion

As can be seen from the foregoing, the new protector constitutes a major improvement over the *99A* protector and over a large range has even greater current-carrying capacity than the line wires with which it is normally associated. These facts indicate the effectiveness of the new device in securing co-ordinated protection in many situations where the *99A* type would be inadequate.

Application of the new protector to joint-use situations involves a number of problems which have not yet been fully resolved. Further investigations are proposed which, it is hoped, will lead to the establishment of suitable practices based upon the solutions and may, in due course, permit the presentation of a fourth paper in this series (in addition to references 1, 2, and this paper) dealing with the subject of co-ordinated protection applications.

Appendix

As explained in the foregoing, the development of the *118A* protector has enlarged considerably the possible field of joint power and communication open-wire facilities as it may be used on many systems whose fault currents or durations are expected to be greater than could be handled by the *99A*. With these increased currents and times, the effect on the fault current of the change-in-impedance of the communication line wire becomes of greater importance than has heretofore been recognized. In brief, the job of the joint-use engineer becomes not only larger, but more complex also.

An effort has been made to devise graphical methods by which the engineering of many joint-use lines can more easily be done. Two examples will be given to show how this method works. The examples are not to be thought of as typical or even as representative of the problems encountered in the engineering of a joint-use line, as each line will be different from all others, and will be a continuously variable thing itself, from the substation to the end of the joint section. The only tools used commonly in each example are the line wire curves presented in Figs. 13 and 14.

Example 1

A 2-wire single-phase multigrounded neutral power system is fed by a 10,000-kva transformer with a phase voltage of 13.8 kv. The power conductor is no. 4 steel-reinforced aluminum cable and protection is furnished by a 100-ampere fuse. The power line is, in part, on joint poles with a pair of *109H* steel telephone open-wire conductors. The *118A* protectors are installed every 1,000 feet on the open-wire pair and the protector ground lead is connected to the neutral conductor.

Assume that a solid contact is made between the power phase wire and one of the telephone wires 8,500 feet from the transformer. The contact occurs midway between two *118A* protectors, but assume that only one of them operates. Other fault-current paths to ground, such as those through station protectors, will be ignored in this example. It is then necessary to know whether the 100-ampere fuse will blow or the line wire fuse.

The total initial impedance of the fault circuit is as follows:

$$\begin{aligned}\text{Power impedance} &= \text{transformer} + \text{power line} + \text{neutral}\\ &= j1.14\omega + 8.5(0.529 + j0.283) + 2.0 + j0\\ &= 6.49 + j3.54 \text{ ohms}\end{aligned}$$

$$\begin{aligned}\text{Telephone impedance} &= 0.5(8.5 + j1.1)\\ &= 4.25 + j0.55 \text{ ohms}\end{aligned}$$

$$\begin{aligned}\text{Total initial impedance} &= 6.49 + j3.54 + 4.25 + j0.55\\ &= 10.74 + j4.09 \text{ ohms}\end{aligned}$$

The initial current for a potential of 8 kv to ground is therefore:

$$I_0 = \frac{8{,}000}{10.74 + j4.09}$$

$$= 696 \text{ amperes}$$

From this initial current, a series of successive approximations on Fig. 13 yields a number of new values of current at later times, caused by the thermal change-in-impedance of the line wire.

These approximate currents and times are as follows:

Current, Amperes	Time, Seconds
696	0
640	0.16
593	0.29
516	0.55
407	1.15
335	2.07
248	4.35

Fig. 19, then, shows the fusing characteristic of *109H* steel line wire, the total clearing time characteristic of the 100-ampere fuse, and the plot of the fault current as a function of time. The conclusion might be drawn from this figure that not only would the 100-ampere fuse fail to blow, but also, the fault current might fall low enough to come to an equilibrium point below the minimum fusing current of the line wire.

The immediately recognizable fallacy in this argument stems from the fact that both the wire and the fuse characteristics were derived from data gathered from constant-current tests and in the simulated example the current is far from constant. Recourse must therefore be made to a graph on which the amounts of energy required to fuse the wire and the fuse link are shown as functions of current. See Fig. 20. In the range of currents of interest, the fusing energy of the line wire is an independent function of current, as shown by the straight line in the figure. The fuse link however has considerable variability in the amount of energy needed to cause it to blow at different currents. At this point it should be admitted that the product of current squared times time is merely proportional to energy, but the problem is greatly simplified if this adjustment will be admitted.

To this figure is then added the curve of energy input to the system as a function of current. This is obtained by graphically integrating the area under a curve on which the square of the fault current is plotted as a function of time. It is now possible to adjust the fuse link curve to take into account the changing fault current. As in all the steps previously outlined, this is an approximate thing, but if sufficient attention is applied, the results should be fairly sound. The adjustment in the curve is achieved by finding first a number of current points at a number of equally spaced points in time, starting at the moment of contact. Time intervals of 0.25 second were used in the example, and the current values were read from Fig. 19. At the initial current point, the fuse-link fusing energy is as shown at the 696-ampere intercept on Fig. 20. At the second current point (610 amperes in the example) the adjusted value of fuse-link fusing energy is approximately equal to the value at current point *1*, plus one half the difference between the values at current points *1* and *2*. Then, the adjusted value at current point *3* is approximately equal to the adjusted value at point *2*, plus one-half the difference between original points *2* and *3*. The remaining points on the adjusted curve are found in the same manner.

It will be seen from Fig. 20, that the 100-ampere fuse will blow and clear when the current has decreased to a value of approximately 360 amperes or at about 1.67 seconds. The time value may be obtained from Fig. 19.

Example 2

A 2-wire single-phase, multigrounded neutral power system is fed by a 10,000-kva transformer with a phase voltage of 13.8 kv. The power conductor is no. 4 hard-drawn copper. Overcurrent protection is furnished by a fuse whose characteristic lies beyond the fusing characteristic of the telephone protector and can therefore be eliminated from consideration in this example. The joint-line wire is *104* copper-steel with *118A* protectors spaced each half mile. The ground electrode of the protectors is connected to the power neutral.

Assume that a solid contact is made at a point midway between *118A* protectors, at a distance of 20,000 feet from the transformer. As before, assume that only one of the adjacent protectors operates and any other current paths to ground will be ignored in this example. Assume also that it is deemed necessary for the line wire to fuse before the protector reaches the "arc transition point," a condition described in the text, or before the protector leads fuse. Fig. 21 shows the characteristics of the line wire and of the protector for the conditions described.

The total initial impedance in the fault circuit is then:

$$\begin{aligned}\text{Power impedance} &= \text{transformer} + \text{line} + \text{neutral}\\ &= j1.14 + 20(0.332 + j0.257) + 4.0 + j0\\ &= 10.64 + j6.28 \text{ ohms}\end{aligned}$$

$$\begin{aligned}\text{Telephone impedance} &= 1.32(2.37 + j0.35)\\ &= 3.13 + j0.46 \text{ ohms}\end{aligned}$$

$$\begin{aligned}\text{Total initial impedance} &= 10.64 + j6.28 + 3.13 + j0.46\\ &= 13.77 + j6.74 \text{ ohms}\end{aligned}$$

The initial current for a potential of 8 kv to ground is therefore:

$$I_0 = \frac{8{,}000}{13.77 + j6.74}$$

$$= 521 \text{ amperes}$$

The fault current as a function of time may again be found by a number of successive approximations on Fig. 14. These are:

Current, Amperes	Time, Seconds
521	0
499	0.76
477	1.68
442	3.30
378	6.75
332	12.25

When these values are plotted in Fig. 21, it would appear that the "arc transition point" would be reached before the line wire fuses. However, the use of the energy relationships and the subsequent adjustments for changing fault current, as shown in Fig. 22, lead to the conclusion that the line wire will fuse before the protector "arc transition point" is reached. This occurs when the fault current has decreased to a value of approximately 360 amperes, or at a time of 8.2 seconds.

It should be pointed out in conclusion that in each of the examples given, the actual co-ordination of power and communication protective equipment was determined. Obviously this cannot be determined on the basis of the time–current characteristics only, as had been done in the past.

References

1. CO-ORDINATED PROTECTION FOR OPEN-WIRE JOINT USE—PRESENT TRENDS. W. R. Bullard, H. E. Weppler. *AIEE Transactions*, vol. 75, pt. I, May 1956, pp. 215–16.

2. CO-ORDINATED PROTECTION FOR OPEN-WIRE JOINT USE—MINNEAPOLIS TESTS. E. G. Albrecht, E. W. Christoferson, A. E. Dietz, J. C. Slothower. *Ibid.*, pp. 217–23.

Vol. 7
Pt. 1
1956

Discussion

M. Fraresso (Hydro-Electric Power Commission of Ontario, Toronto, Ont., Canada): From a power utility point of view the final objectives of joint use of poles is the simplification and standardization of procedures which will result in greater economics without loss of safety and reliability. The development of the new protector has greatly enhanced the possibility of securing these objectives but now further evaluation of co-ordination protection is required to take full advantage of the recent developments. It is now possible to secure co-ordination of the power protective devices with the more commonly used communication line wires, provided line wire burndown removes the power potential from the communication circuit. Example 1, Appendix, indicates that this co-ordination is obtained even when it is assumed only one *118A* protector operates for a power cross in the middle of the joint use section.

The authors are to be congratulated on their presentation of a timely subject, and the work that has been accomplished to date. They have mentioned the possibility of employing the principle of "burn-down providing satisfactory co-ordination." This revolves around the problems of ensuring that the burn-down occurs in the vicinity of the power cross hence securing de-energization of the communication circuits, as yet no general conclusions being established. It would appear desirable to give further consideration to this approach as it may provide a simple solution to a complex problem.

It might be well to speculate some of the reasons whereby burn-down could occur, subject to confirmation by actual tests. These reasons are:

1. Contact area will be small, resulting in a relatively high contact resistance compared with the resistance of communication wires.
2. Oxides on the surface of the conductors will tend to increase the contact resistance causing localized arcing at the point of contact, and high arc temperatures in the neighborhood of 6,500 degrees Fahrenheit could melt the communication conductor at the point of contact. This situation would be further enhanced at high current values because of the spattering effect of metal (based on the welding characteristics of 1/8-inch rods when the arcing current exceeds 200 amperes).
3. The effect of movement of the power wire across the communication wire will result in deformation and pitting of the communication conductors and increasing the possibility of burn-down at the point of contact.
4. Provided power cross occurs between two joint-use protectors, the second protector will operate after the first has burned clear. This will subject the point of contact to prolonged arcing.

If burn-down becomes accepted as a criterion of co-ordination, considerable extension of joint use with the new protector could be obtained especially if the low- and high-current ends of the present protector damage curve could be improved to ensure co-ordination with *104* copper steel line wire over the whole range of currents.

For joint use near power substations, the *118A* protectors would be required at intervals less than the usual spacing of 1/2 mile in order to limit plant rise to 3,000 volts. With the added restriction of increased impedance of communication wire as a function of current and time, as illustrated in the paper, it may become necessary further to reduce spacing of *118A* protectors drastically to ensure that the maximum plant rise of 3,000 volts would not be exceeded. It would therefore be desirable to remove this restriction on plant rise by adopting an improved fuseless subscribers' protector and relying on co-ordination with the drop wire time–current fusing characteristics similar to the co-ordination obtained for cable protection.

Where the drop wire can be used for co-ordination, if a power cross occurs at a point having several subscribers located between it and the *118A* protector, the communication conductor between the power cross and the protector will be subject to an additional 10 per cent of the I^2t equivalent heating per subscriber (for *109* steel) than the remainder of the joint use section. This is based upon the I^2t equivalent thermal capacity of the drop wire compared to that of the line wire.

If it were possible to remove the restriction on plant rise by allowing communication line wire burndown to constitute co-ordination, a significant step would be taken in simplifying joint use engineering requirements without loss of safety and reliability.

H. M. Ellis, J. W. Phelps, C. L. Roach, and **R. E. Treen:** It is safe to say that the objectives of both power and telephone companies as regards the safety and reliability of their separate functions are very much the same. In both utilities the objective is ever-increasing service to more and more customers at a constantly reducing cost with a consistent margin of safety. Naturally, an extension of open-wire joint use is one method whereby considerable savings can be achieved. The introduction of the *118A* protector does much for the extension of joint use, as it is now possible, under present practice, to engineer joint-use lines up to the fusing point of the communication line wire, taking into account, of course, the effect of the changing line wire impedance under fault current conditions, as illustrated briefly in the paper.

There are, however, a large number of power distribution systems that must use protective apparatus whose time–current characteristics are greater than the fusing characteristics of communication line wire. If these systems are to be constructed jointly with telephone plant, interruption of any fault current flowing on the telephone conductors must be accomplished through burn-down of the line wire. The reasons listed by Mr. Fraresso in his discussion would seem to assure the safe burn-down of telephone plant, and the same lines of thought were followed prior to the tests at Minneapolis (see reference 2 of the paper). However, a limited number of tests there indicate otherwise. The results of these few tests, which were not reported in the paper because of their inconclusive nature, have been rationalized as follows:

1. The high thermal conductivity at the contact point afforded by the usually heavier power conductor causes heat to be quickly conducted away from the point of contact. This effect appears to be greater than the combined effect of arcing and high contact resistance at the contact point. In the line-wire fusing tests where the power wire was allowed to drop on the line wire, the line wire did not fuse within perhaps an inch of the contact point. It did, however, become hot enough in several cases to become welded to the power conductor. The sequence of events might have been this: The heat of the initial arc at the contact and of the concentrated current through that small portion of the wire caused that part of the wire to become plastic. The weight of the power wire then caused it to be imbedded in the line wire. Finally, the heavier power wire conducted the heat away from the contact point, causing the two wires to be welded together.
2. If the telephone line wire has been nicked by wire pullers, pliers, etc., not however to the extent that its strength is appreciably reduced, there will be localized heating at the constricted point which elevates the temperature at that point faster than the rest of the wire and causes the wire to fuse at that point only. This latter effect was so pronounced that considerable care had to be taken in the placing of the wire for test so that it would fuse uniformly.

Since the few tests performed at Minneapolis indicate the possibility of the welding of the two conductors at the contact point, or the fusing of the line wire at some remote point, there may be considerable effort necessary to investigate the effects of power contacts on typical field installations. In addition to these technical points, there are further intangible questions such as the telephone customer reaction to the service outage.

Assume next that a satisfactory solution is found to the foregoing problems so that there need be no worry about the actual physical separation of power and telephone conductors. There must then be assurance that the fault current will not flow for such a period of time that the increased resistance of the line wire causes the fault current to drop below the minimum protector operating or line wire fusing current. In the second example listed in the paper, it is apparent that if more telephone wire had been in the circuit, the initial current would have been lower, the rate of change of current with time would have been more pronounced, and the end result would have been the destruction of the protector before the fusing of the line wire. The obvious conclusion to be drawn is that unless some engineering effort is put into the layout of the telephone protection system, the system must necessarily be conservatively designed. In this case, the cost of the protection may well offset the savings in engineering effort.

It should further be pointed out that the graphical determination of the results in each of the two examples in the paper is an approximation and tests are necessary to prove the feasibility of the graphical method. Such tests would also more closely define the protector capability at the lower currents to show if further modification is necessary.

On the question of a fuseless protector for higher voltage joint-use lines, it was considered in 1950 that such would be feasible at rural telephone stations which were also served by multigrounded neutral power systems and where power and telephone were commonly grounded. Active development effort on the problem, including the design of the protector, was begun 2 years ago and is still continuing. It is safe to say that the protector can be made at increased cost over the fuseless protector now used on stations served by metallic sheathed cables. In either case, there will still be a fusible element in the subscriber's circuit, namely, the terminal cable conductors and bridle wires.

The administration of fuseless protectors at stations on open-wire lines may yet prove to be a greater problem than the design of the device. This is due to the fact that careful attention must be paid to the ground to which the protector is connected. This is important for two reasons: The magnitude of the fault current that flows through the station protector is in part determined by the resistance of the station ground, and the potential difference between the power and telephone networks in the subscriber's home is proportional to the impedance between the two grounds. In the ideal case both utilities are grounded to an extensive public water-piping system with very low impedance to remote earth. A contact, therefore, causes considerable current to flow into the ground, blowing the station fuse or the fusible element in the drop wire run. The current, however, does not cause a dangerous rise in potential at the station with respect to remote earth, neither does it cause potential differences to exist between the power and telephone systems in the home. Unfortunately, many of the stations served by open wire do not have water systems. Bonding of the telephone ground electrode to the multigrounded neutral ground must then be resorted to in order to minimize potentials between the power and telephone wiring in the station.

A Magnetic Amplifier Switching Matrix

D. KATZ
ASSOCIATE MEMBER AIEE

By using conventional high gain or bistate magnetic amplifiers as building blocks, switching networks may be derived which perform many computer or control functions. One such network which has been built and tested is a 4-digit binary-to-decimal translator. This translator was designed to provide a display in decimal numbers of binary-coded information. Basically a switching matrix made up of simple self-saturating magnetic amplifiers, this translator illustrates both the matrix technique and an advantageous application. The translator matrix operation, the specific bistate magnetic amplifier building blocks used, the decimal displays obtained, and their individual and combined characteristics are discussed in this paper.

Although magnetic core switching array applications in pulse circuits have been discussed in the literature,[1–4] the use of magnetic amplifiers to switch continuous power (rather than pulse power) has received little attention. In general, a bistate magnetic amplifier may be viewed as a contactless switch having some functional similarities to a relay. Without any moving parts, such a switch electrically isolates inputs and outputs, can provide high power amplification, and can be electrically set or biased to switch at a precise net signal value appearing in several control windings. In a manner similar to the application of magnetic cores in pulse-switching arrays, such bistate magnetic amplifiers may also be used in switching matrices. The matrix form and the building block magnetic amplifier used depend, naturally, upon the function performed, the inputs used, and the desired outputs. The specific matrix discussed in this report, a binary-to-decimal translator, acts, in effect, as a multiposition switch which distributes energy to any one of a number of positions in response to signal information. The translator accepts 4-digit binary-coded numbers on four leads and a common return, and it performs such translation and amplification that an easily read decimal number display is obtained from the 10-lead output with the use of either a 10-element neon gas tube or 10 filamentary lamps.

The translation problem arises from the need to present digital computer data in an illuminated decimal number display. With the binary computer involved, the number selected for display is converted to parallel binary-coded decimal so that each decimal digit of this number appears on four leads as a 4-digit binary-coded word. For each decimal digit there is then required a further translation from the binary-coded decimal to one out of 10 decimal signals with which to light indicating lamps or devices. This is accomplished by a magnetic amplifier-switching matrix or translator. The 4-digit binary-to-decimal conversion table in Fig. 1 shows the binary inputs and the required translation. As obtained from the computer logic packages, a binary "1" signal has an average value of 1.4 volts. The binary "0" is the absence of any signal or zero volts.

The operation of the magnetic-amplifier translator matrix to select one of 10 outputs in response to the binary inputs is described in the following section. Two types of magnetic-amplifier building blocks which have been used in such a matrix are discussed in the section "Bistate Magnetic Amplifier Switches." Discussion of decimal displays and translator characteristics, and the conclusions follow in that order.

Magnetic Amplifier Translator Matrix

For each decimal digit in the display, the associated binary to decimal translator has 10 possible 4-digit binary inputs as shown in the table of Fig. 1. With each binary input a unique one of 10 bistate magnetic amplifiers in a matrix is triggered ON while the others are held OFF. Only the fired (or ON) bistate unit transfers electrical power to its output lead so that for each binary input only one of 10 output leads is energized. The energized lead then causes a decimal number to be displayed. The schematic of such a matrix is shown in Fig. 2.

Fig. 2 shows 10 bistate magnetic amplifiers with their control windings arranged in a 4-by-10 matrix. The eliminated control windings, which would have filled the matrix, are not needed because the selection of the matrix is one of 10 inputs rather than one of 16 bits available from a 4-digit binary word. In this instance, half-wave magnetic amplifiers consisting of one core with a power winding, several control windings, and a bias winding are shown as the bistate units of the matrix. The dots indicate winding polarities, with current which enters the winding through a dotted end called positive. Thus, with the indicated bias current direction and bias winding polari-

---

Paper 56-205, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 24, 1955; made available for printing December 15, 1955.

D. Katz is with Bell Telephone Laboratories, Inc., Whippany, N. J.

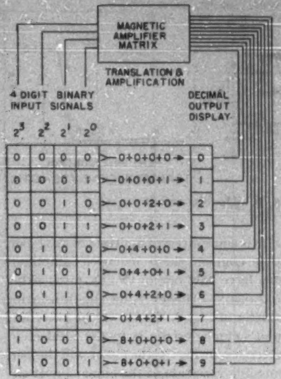

MAGNETIC AMPLIFIER MATRIX
TRANSLATION & AMPLIFICATION
4 DIGIT BINARY INPUT SIGNALS
DECIMAL OUTPUT DISPLAY
2³ 2² 2¹ 2⁰
0 0 0 0 ▷ 0+0+0+0 → 0
0 0 0 1 ▷ 0+0+0+1 → 1
0 0 1 0 ▷ 0+0+2+0 → 2
0 0 1 1 ▷ 0+0+2+1 → 3
0 1 0 0 ▷ 0+4+0+0 → 4
0 1 0 1 ▷ 0+4+0+1 → 5
0 1 1 0 ▷ 0+4+2+0 → 6
0 1 1 1 ▷ 0+4+2+1 → 7
1 0 0 0 ▷ 8+0+0+0 → 8
1 0 0 1 ▷ 8+0+0+1 → 9

Fig. 1. Binary coded decimal-to-decimal translation

ties, all the biased units have negative ampere turns bias. The bias windings are weighted by having N_B, $2N_B$, or $3N_B$ turns as indicated in Fig. 2. All control windings have equal turns (N_C).

To explain the operation of the matrix, the bistate units may be assumed to have an idealized transfer characteristic as shown in Fig. 3. The output state (ON or OFF) is determined by the net control or the algebraic sum of the ampere-turns of all the control and bias windings on a bistate unit. Any binary digit 1 input causes an average control current I_c to flow through the control windings of the associated column in the direction indicated in Fig. 2. Therefore, on a per-unit basis with N_cI_c as a control unit, a binary 1 signal in a control winding is a ± 1 control where the sign is determined by the polarity of the control winding. With a binary 0 input there is no current flow. The bias current (I_B) is set so that $|N_BI_B| = |N_cI_c|$, or on a per-unit basis, the bistate units having N_B turns are biased to -1, those with $2N_B$ turns to -2, and those with $3N_B$ turns to -3. As seen in the transfer curve of Fig. 3, any unit with zero or greater net control is fully ON. Any unit with -1 or less net control is fully OFF. Now, observing the bias values and the control winding polarities, the table of Fig. 4 can be constructed showing the net control (bias plus control input) on every bistate unit for each of the 10 binary inputs. For example, with a (0, 0, 0, 0) binary input, all the bistate units will have some bias value (-1, -2, or -3) except for the decimal 0 unit which has no bias. Consequently, only the decimal 0 unit, having zero net control, is ON. For any other input the 0 unit is driven off because all of its control windings have negative polarities. For a (0, 0, 1, 1) binary input, only the decimal 3 unit will have zero control while all others are negative. For instance, unit 7 will have a bias of -3, a $+1$ control in the 2^0 column, and a $+1$ in the 2^1 column, yielding a net control of -1. The table, Fig. 4, which tabulates all these values, clearly shows that for each of the 10 binary inputs, a unique one of the 10 bistate devices is driven on.

The particular switching matrix described switches electric power suitable for directly operating the output displays. The two output states of the bistate units here employed are a continuous voltage or practically no voltage output as derived from the a-c power supply. In the pulse circuit magnetic core matrices, which are more widely discussed in the

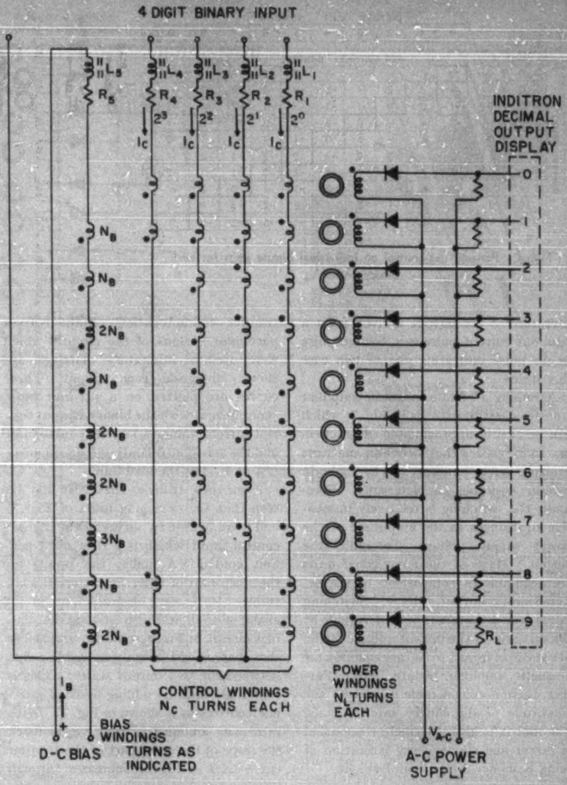

4 DIGIT BINARY INPUT
L₅
R₅
L₄
R₄
L₃
R₃
L₂
R₂
L₁
R₁
2³
2²
2¹
2⁰
I_C
I_C
I_C
I_C
INDITRON DECIMAL OUTPUT DISPLAY
0
1
2
3
4
5
6
7
8
9
N_B
N_B
2N_B
N_B
2N_B
2N_B
3N_B
N_B
2N_B
R_L
I_B
+
CONTROL WINDINGS N_C TURNS EACH
POWER WINDINGS N_L TURNS EACH
BIAS WINDINGS TURNS AS INDICATED
D-C BIAS
V_A-C
A-C POWER SUPPLY

Fig. 2. Magnetic-amplifier translator matrix

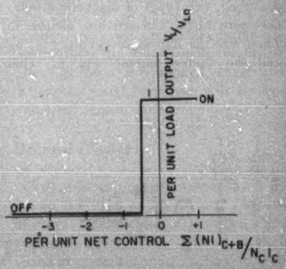


Fig. 3. Idealized bistate magnetic-amplifier transfer characteristic

BINARY INPUTS 2^3	2^2	2^1	2^0	BISTATE UNIT NUMBER 0	1	2	3	4	5	6	7	8	9
				PER UNIT NET CONTROL									
0	0	0	0	0	-1	-1	-2	-1	-2	-2	-3	-1	-2
0	0	0	1	-1	0	-2	-1	-2	-1	-3	-2	-2	-1
0	0	1	0	-1	-2	0	-1	-2	-3	-1	-2	-1	-2
0	0	1	1	-2	-1	-1	0	-3	-2	-2	-1	-2	-1
0	1	0	0	-1	-2	-2	-3	0	-1	-1	-2	-1	-2
0	1	0	1	-2	-1	-3	-2	-1	0	-2	-1	-2	-1
0	1	1	0	-2	-3	-1	-2	-1	-2	0	-1	-1	-2
0	1	1	1	-3	-2	-2	-1	-2	-1	-1	0	-2	-1
1	0	0	0	-1	-2	-1	-2	-1	-2	-2	-3	0	-1
1	0	0	1	-2	-1	-2	-1	-2	-1	-3	-2	-1	0

Fig. 4. Per-unit net control on individual bistate units for each binary input

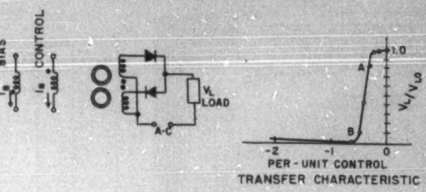


Fig. 6. Doubler circuit

literature[1-4] the output derived from a read-out current pulse is a voltage pulse or no pulse indicating one of two core flux states.

Assuming a bistate magnetic-amplifier transfer characteristic as in Fig. 3, which can be closely approximated in practice, one soon realizes that switching matrices utilizing such building blocks embody several desirable characteristics. Foremost, the switching is relatively insensitive to changes in the amplitude of the power supply voltage. Secondly, the output voltage of the nonselected units are negligible in comparison to the selected unit's output voltage. Switching ratios (the ratio of the output voltage at the ON state to the output voltage of the OFF state) of from 5 to 500 are common for magnetic amplifier bistate units. Further, there is considerable margin for the amplitude of the binary input signals. The matrix is also extremely economical of power and shows every indication of being economical in size and weight.

Bistate Magnetic Amplifier Switches

Recent papers and articles[5-7] both here and abroad about bistate magnetic amplifiers attest to the increased interest in such devices. In the translator matrix described, two types of magnetic amplifiers were used as the switching elements. These are the half-wave circuit, Fig. 5, and the doubler circuit, Fig. 6. The transfer characteristics, obtained from particular designs of the circuits which were used in laboratory matrices, are shown alongside each circuit. These curves are plotted on a per-unit basis (normalized) with the binary 1 input control current value (I_c) as the control unit and the zero control output load quantity (I_{L0} or V_{L0}) as the load unit.

From these transfer curves, it may be seen that the switching units of Figs. 5 and 6 are bistate by virtue of the digital-control input which is a 0 or $a \pm 1$ per-unit control. Actually, the points on the steep part of these characteristics are stable. If it were desirable, as in an analogue-to-digital conversion, the doubler circuit of Fig. 6 could be designed so that there would be truly no stable points between the two output states. This is commonly done by adding positive external feedback as is shown in Fig. 7. With increasing amounts of positive feedback the slope of the steep part of the transfer curve (AB in Fig. 6) increases through infinity and becomes negative, i.e., the curve theoretically bends back upon itself. However, since the points of infinite or negative slope are unstable, the transfer characteristic takes on the form shown in Fig. 7.

With a digital-control input, high-gain magnetic amplifiers, such as those of Figs. 5 and 6, make very satisfactory switching elements. Since a control signal is either a 1 or 0 on a per-unit basis and the bias is some integral multiple of the 1-per-unit control, these magnetic amplifiers can have only one of two outputs, either OFF or ON. However, for the translating matrix of Fig. 2, the closer the switching unit's characteristic is to the ideal characteristic of Fig. 3, the closer will the permissible variation of the binary input signals approach a theoretical maximum of ±20.8 per cent (%).

The half-wave circuit of Fig. 5 uses only one core and one diode, but necessitates a high impedance to alternating current in every control and bias loop in order to minimize the transformer action between load and control meshes. In the steep region of the transfer curve, AB in Fig. 5, the gain or slope varies inversely with the induced a-c components in the control meshes. Consequently, in the translating matrix, Fig. 2, which uses half-wave magnetic amplifiers, there is required an inductor, L_1, L_2, L_3, L_4, and L_5, in each control mesh. Series resistors, R_1, R_2, R_3, R_4, and R_5, are padding resistors to compensate for winding resistance variations and to bring the loop resistance up to a desired value. The output of the half-wave circuit is, of course, half-wave recitfied alternating current. Since the Inditron gas tube display requires a direct voltage on the selected element, and since it was desirable to minimize the number of cores and diodes, the half-wave circuit was used in the translating matrix with a gas tube decimal display as shown in Fig. 2.

A balanced 2-core magnetic amplifier like the doubler circuit of Fig. 6 permits the use of low-impedance control meshes

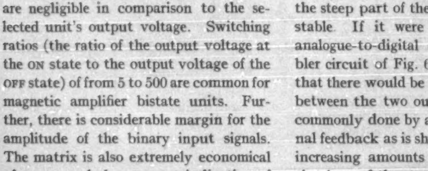


Fig. 5. Half-wave circuit

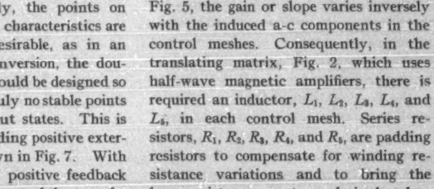


Fig. 7. Doubler circuit with external positive feedback

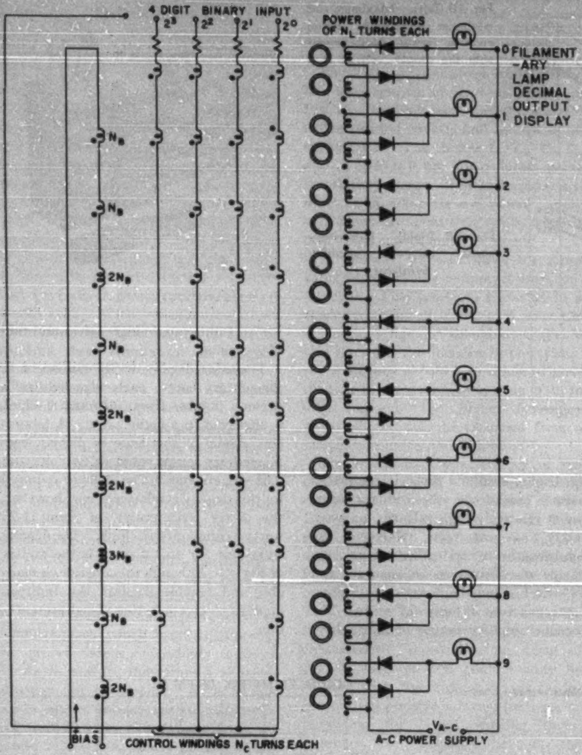


Fig. 8. Translator matrix

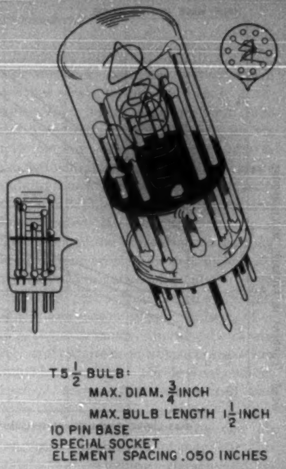


Fig. 9. National Union G1-10 Inditron neon decimal display tube

because the induced fundamental and odd harmonics of the a-c supply are bucked out in the control loops. Physically, this magnetic amplifier has two tape-wound cores each having an a-c power winding on it. These are then stacked and common control and bias windings are wound around the two stacked toroids. The output of the doubler circuit is alternating current. To provide a d-c output, two more diodes would be required for a bridge connection in the output circuit. However, for filamentary lamps the a-c output is satisfactory and the doubler circuit of Fig. 6 was used in a translator matrix with a lamp display as is shown in Fig. 8. The translator matrices of Figs. 2 and 8 are the same in principle and differ only in the type of magnetic-amplifier switch circuit used.

The doubler circuit eliminates the need for chokes in the control loops, does not severely restrict the number of control winding turns, and lends itself to the incorporation of positive external feedback. The last two factors permit very high-gain or true bistate magnetic amplifier operation.

The switching ratio of both the magnetic-amplifier circuits discussed here is a function of the number of power turns and the load impedance, for these largely determine the OFF-state output voltage. The switching ratio of the half-wave circuit, operating into a high-impedance load (51,000 ohms) in the translator, is close to six with the design values used. This ratio is more than adequate because the gas display tube's high breakdown and extinction potentials act to discriminate between ON and OFF switch output states. The doubler circuit, operating into a low-impedance filamentary lamp load (100 to 700 ohms), has a switching ratio in the order of 200. The filamentary lamp in the display requires a switching ratio of about 15.

The magnetic amplifier switches discussed are closely akin to magnetic-core switches used in pulse circuits. The former, however, use a continuous a-c power supply or drive to obtain a continuous ON or a zero OFF electrical output for the switch states, while the latter use a single-pulse drive to obtain a pulse or no-pulse indication of the switch states. Both types of switching elements have exceptionally long life, are reliable, and have large-power handling ability. The switching magnetic-amplifier's speed of response is, of course, limited to one or more cycles of the a-c power supply frequency. Although practically usable upper limits for the power supply frequency have not been investigated, there are indications that difficulties will be encountered at over 50 kc.

Decimal Displays

Two different devices were selected to obtain the illuminated decimal number display. One of these, a gas tube, such as the National Union Inditron, is preferred at present because of the clarity of the display obtained. This 10-element neon tube, with each element formed into

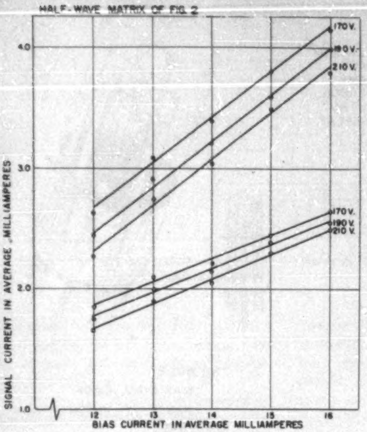


Fig. 10 (left). Maximum and minimum permissible signal currents versus bias current with the supply voltage as a parameter

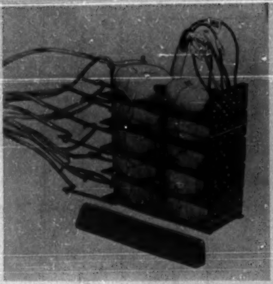


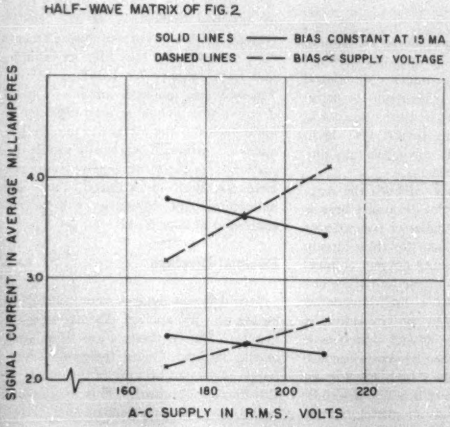
Fig. 12 (right). Laboratory model magnetic amplifier translator

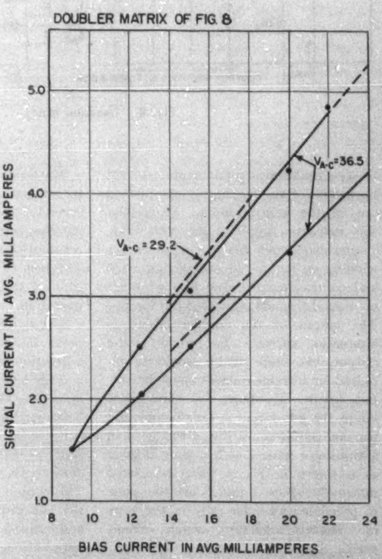
a decimal number from 0 through 9, is pictured in Fig. 9. With the half-wave rectified output of the half-wave translator matrix, the selected element is held about 125 volts rms more negative than the other elements. The selected element becomes a glowing cathode, the other elements act as shared anodes, and a well-defined, easily read decimal number display results. Since it requires a gaseous breakdown, the Inditron operates at relatively high voltages, approximately 250 peak volts. However, the high-breakdown and extinction voltages provide discrimination between ON and OFF switches and the power consumption of the gas tube is only 0.2 watt. The magnetic-amplifier matrix of Fig. 2 driving an Inditron provides a clear and reliable display.

The other display device consists of 10 filamentary lamps, each edge lighting a decimal number from 0 through 9 which is engraved in a lucite plate. A laboratory model of this type of display uses filamentary lamps rated at 28 volts and 0.04 ampere rms. The voltage required for this display is relatively low; however, the power consumption of about 1.25 watts is comparatively high. The driving translator for this display is the matrix of Fig. 8. Although this provides a readable and reliable display, the Inditron appears to give a better visual presentation.

Fig. 11 (below). Maximum and minimum signal currents versus a-c supply voltage

Fig. 13 (right). Maximum and minimum signal currents versus bias current





Translator Characteristics

The translator matrices of Figs. 2 and 8 using half-wave and doubler building blocks, have been built and tested. Since the Inditron display was preferred, a mock-up of a 5-digit decimal display with five half-wave translator matrices and five gas display tubes was assembled and demonstrated. For comparative purposes, one model of the doubler translator matrix has been built and operated. The specific designs of both types of matrices for which characteristics are given here operate with the same inputs. The binary 1 is 1.4 average volts and the binary 0 is zero volts. With a maximum current drain from the signal source of 4 milliamperes, these translators are designed for a nominal binary 1 signal current of 3 milliamperes average. The a-c power supply is 400 cycles per second.

Typical operating characteristics of a half-wave translator matrix with an Inditron tube are shown in the curves of Figs. 10 and 11. From the curves of maximum and minimum signal currents versus bias current (Fig. 10), it is seen that at 15-milliampere average bias and with a supply voltage of 190 volts rms, the binary 1 input signal current may vary between 2.38 and 3.62 average milliamperes. For a nominal binary 1 input signal current of 3 milliamperes, which is set by choosing the proper values of padding resistors R_1, R_2, R_3, and R_4, this means a ±20.6% tolerance for the inputs. With an a-c power supply voltage variation between 170 and 210 volts, this signal margin becomes ±15%. A clearer picture of the translator's insensitivity to a-c voltage amplitude is given in the plot of permissible signal currents versus the alternating voltage, Fig. 11. The solid lines are for a constant 15-milliampere bias and the dashed lines are for bias proportional to the a-c supply voltage.

Other pertinent electrical characteristics of this translator may be listed as follows:

1. Frequency—A 5% variation in the frequency (380 to 420 cps) of the supply voltage has a negligible influence on operation.

2. Power consumption—The a-c power input required is less than 1/2 watt per translator.

3. Temperature—Temperature compensation over a range of −50 to +60 degrees centigrade can be made implicit in the design, for the temperature change in bias current can compensate for the change in signal currents.

4. Response Time—Switching from one decimal digit to another takes about 0.01 second with a 400-cycle power supply.

Physically, a laboratory model of the half-wave matrix appears as shown in Fig. 12. In production such units would utilize printed circuits and be cast in an epoxy resin. The chokes L_1, L_2, L_3, L_4, and L_5 of Fig. 2, are not included in the translator package. The packaged translator for a 400-cycle a-c power supply would measure about $1^7/_{16}$ by 3 by 3 inches and weigh approximately 1 pound.

At present, an equivalent translator using a 4,000-cycle a-c supply is being constructed. This promises to result in a space reduction of about 40% over the 400-cycle design. Although unsought in this instance, this increase in power supply frequency should also result in a reduction in switching time from 0.01 to 0.001 second. The higher frequency power source may be obtained from a magnetic-coupled multivibrator.[8,9]

The operating characteristics of the doubler matrix with a filamentary lamp display are shown by the curves of Fig. 13. At a 16-milliampere average bias, with a 36.5-volt rms 400-cycle supply, and again considering 3 milliamperes average as the nominal signal, the signal margin is slightly more than ±10%. This could be increased by increasing the switching unit's gain so that its transfer characteristic approaches the ideal of Fig. 3. Greater gain could readily be realized with the doubler circuit by incorporating positive external feedback or increasing the signal winding turns. However, the difference between the signal margins of the half-wave and doubler matrices (±20.6% versus ±10%) is not due to any difference in transfer characteristics but to a-c induced effects in the control loops of the half-wave matrix which beneficially increase its margins. However, these induced effects, which are limited by the chokes, L_1, L_2, L_3, L_4 and L_5, must not be allowed to become too large, for there would be a concomitant loss in gain in each of the half-wave switching units. When this loss in gain is sufficient to move the point B on the transfer characteristic of Fig. 5 to the left of the -1 per-unit control, the translating matrix does not function properly.

The doubler matrix requires an a-c input power of 2 watts with the filamentary lamp using about 1.25 watts of this. Such a matrix would package in much the same manner and size as the half-wave matrix and would not require control loop chokes.

Conclusions

Basically, this report described a magnetic-amplifier switching matrix which switches power to one of 10 leads in response to a binary input on four leads and a common return. Such a matrix may employ any of several suitable magnetic-amplifier circuits as the bistate building block. Two circuits, the half-wave and doubler circuits, were used and demonstrated in laboratory matrices. The output of the translating matrix may be used to power devices such as indicating lamps or a gas display tube. These translators permit wide margins for input signals and power supply voltage, offer large power economy, are rugged and reliable, and require no more space than other equivalent devices. The decimal number displays derived are well defined and easily read.

This translator demonstrates one advantageous application of magnetic amplifiers as switching elements. Certainly, in switching circuits where switching times of less than 1 millisecond are not required and where continuous rather than pulse power is to be switched, bistate magnetic amplifier switches have much to offer and present such broad possibilities that their use for other applications in place of more conventional devices appears well worth consideration.

References

1. Static Magnetic Matrix Memory and Switching Circuits, **Jan A. Rajchman.** *RCA Review*, Radio Corporation of America, Princeton, N. J., June 1951, pp. 183–201.

2. A Magnetic Matrix Switch and Its Incorporation Into a Coincident-Current Memory, **Kenneth H. Olsen.** *Report R-211*, MIT Digital Computer Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., June 1952.

3. Logical and Control Functions Performed with Magnetic Cores, **S. Guterman, R. D. Kodis, S. Ruhman.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, no 3, Mar. 1955, pp. 291–98.

4. Pulse Switching Circuits Using Magnetic Cores, **M. Karnaugh.** *Ibid.*, vol. 43, no. 5, May 1955, pp. 570–83.

5. Magnetic Amplifiers in Bistable Operation, **L. A. Finzi, G. C. Feth.** *AIEE Transactions*, vol. 74, pt. I, Nov. 1955, pp. 592–98.

6. The Single-Core Magnetic Amplifier as a Computer Element, **R. A. Ramey.** *Ibid.*, vol. 71, pt. I, 1952, pp. 442–46.

7. Transient Processes in Contactless Magnetic Relays, **I. B. Negnevitzkyi, L. L. Samurina.** (Translation from Russian.) *Engineers' Digest*, New York, N. Y., Apr. 1954.

8. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, **G. H. Royer.** *AIEE Transactions*, vol. 74, pt. I, July 1955, pp. 322–26.

9. A Variable Frequency Magnetic-Coupled Multivibrator, **R. L. Van Allen.** *Ibid.*, pp. 356–61.

Measurements on High-Speed Magnetic Servo Amplifiers with 2-Phase Motor Load

W. A. GEYGER
MEMBER AIEE

Synopsis: Measurement of steady-state characteristics and transient-response time of high-speed magnetic servo amplifiers with 2-phase motor load offers two special problems:

1. The output current of a magnetic amplifier with rectangular-hysteresis-loop core material represents a peculiarly truncated sine wave which consists of a fundamental (60- or 400-cycle) component and an array of harmonics having relatively large magnitudes. The 2-phase motor, however, responds only to the fundamental component of this output current multiplied by the sine of the actual phase angle between it and the substantially sinusoidal line-field current of the motor. Thus, in this case, conventional methods for measuring output-input characteristics by means of a separately excited wattmeter are not always applicable.

2. The transient-response time of modern forms of magnetic amplifiers is in the order of 1 to 4 half-cycles of the power-supply frequency. Therefore, for studying the dynamic properties of such high-speed amplifiers it will be advisable to employ a transient analyzer which produces a synchronous square-wave signal voltage and permits, also with higher power-supply frequencies, direct visual observation of a stationary pattern on a dual-beam oscilloscope.

This paper is concerned with the solution of these two problems. At first, it describes the basic principle and the advantages of the "torque-balance method" which permits direct measurement of magnetic-amplifier output in terms of the "equivalent sinusoidal output voltage" for stalled-motor operating conditions. Then, it illustrates the use of a novel form of square-wave signal-voltage transient analyzer with a "magnetic switch" which consists essentially of a special saturable-reactor circuit producing a synchronous unidirectional current with nearly rectangular waveshape. When applying this analyzer, a dual-beam oscilloscope displays the wave forms of input and output voltages of the magnetic amplifier under test simultaneously on a common time base in such a way that these wave forms are exhibited on a long-persistence screen as a stationary pattern. Thus, it is possible to study the effect of various circuit-parameter changes on actual speed of response and to determine optimum design. Finally, this paper gives results of additional measurements on the circuit under test. Measurement of bandwidth on a 400-cycle position servomechanism illustrate the dynamic performance of the circuit when operated in connection with a phase-sensitive rectifier in a closed-loop servo system. Measurements concerning reduction of asymmetry zero-drift errors make it evident that such errors are reduced to about 1/40 of their original values as a result of the application of a large amount of electric negative feedback resistance-coupled into the control-circuit loop.

DURING the past 10 years, a considerable part of the research and development work on magnetic amplifiers has been conducted toward the application of such amplifiers for the control of small 2-phase induction-type reversible motors. These motors are extensively used in various kinds of servomechanisms because of their reliability, rapid response, and absence of commutator sparking. Magnetic servo amplifiers, based upon various forms of full-wave or half-wave push-pull circuits, have been successfully applied in the realm of instrumentation and automatic control; and they have proved to be particularly suitable for use in high-performance instrument-type servos for remote positioning applications. Another interesting possibility consists in employing such amplifiers in the operation of high-speed self-balancing potentiometers and bridge networks.

The designer, when investigating steady-state characteristics and transient-response properties of high-speed magnetic servo amplifiers with 2-phase motor load, faces two special problems. At first, he must consider the fact that the output current of a magnetic amplifier with rectangular-hysteresis-loop core material is, in general, highly nonsinusoidal. It represents a peculiarly truncated sine wave which consists of a fundamental (e.g., 60- or 400-cycle) component and an array of harmonics having relatively large magnitudes. The separately excited 2-phase motor, however, responds only to to the fundamental component of this output current multiplied by the sine of the actual phase angle between it and the substantially sinusoidal line-field current of the motor.

An obvious method for measuring this fundamental component consists in applying a single-phase wattmeter which is separately excited by means of a sinusoidal auxiliary current of fundamental frequency. This method, however, does not consider those effects on the operation of the motor which are caused by phase-angle variations in the magnetic-amplifier output circuit. It is, therefore, desirable to provide another method, the basic principle of which includes the actual effects of these phase-angle variations on the operation of the 2-phase motor.

The second problem resides in the fact that the transient-response time of modern forms of magnetic amplifiers is in the order of 1 to 4 half-cycles of the power-supply frequency. For this reason it is advisable to employ a transient analyzer which produces a synchronous square-wave signal voltage and permits direct visual observation of a stationary pattern on a dual-beam oscilloscope. However, application of the simple magnetic-amplifier transient analyzer employing a synchronously operated mechanical switch[1] is limited to low power-supply frequencies where a mechanical switch can be operated in a reliable way. It is, therefore, necessary to provide another arrangement which may be used equally well in the range of higher power-supply frequencies, for example those derived from magnetic frequency multipliers.[2]

This paper describes measurement of steady-state characteristics and transient-response time of a typical single-stage design of a high-speed magnetic servo amplifier with 2-phase motor load. It suggests application of the torque-balance method which permits direct measurement of magnetic amplifier output in terms of the equivalent sinusoidal output voltage for stalled-motor operating conditions. Then, it illustrates the use of a novel form of square-wave signal-voltage transient analyzer with a magnetic switch which consists essentially of a special saturable-reactor circuit producing a synchronous unidirectional current with nearly rectangular waveshape. This arrangement

---

Paper 56-218, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–Feburary 3, 1956. Manuscript submitted September 27, 1955; made available for printing December 1, 1955.

W. A. GEYGER is with the U. S. Naval Ordnance Laboratory, Silver Spring, Md.

Acknowledgment is made to A. D. Krall and E. T. Hooper of the Naval Ordnance Laboratory for their work in mathematically analyzing the self-balancing magnetic amplifier.[27]

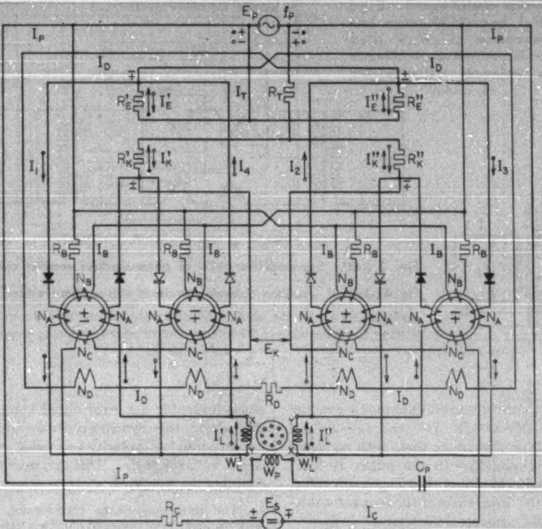


Fig. 1. Circuit diagram of the self-balancing half-cycle-response magnetic servo amplifier with separately excited 2-phase motor

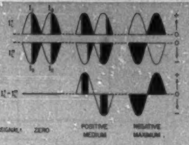


Fig. 2. Variations of the average values of the four half-cycle pulses, I_1 to I_4, and wave forms of the resultant motor currents I_L', I_L'', and $I_L'-I_L''$ for the typical example that, with no-signal conditions, the firing angle of the four saturable-reactor elements is 90 degrees

has proved to be a valuable tool for experimental studies on the dynamic properties of high-speed magnetic amplifiers, particularly those having full response to a sudden change of input signal within the time of 1 half-cycle of the power-supply frequency.

The Circuit Under Test

Although the measurements described in this paper are basically applicable to various types of high-speed magnetic servo amplifiers with 2-phase motor load, it is necessary to describe one representative form of such an arrangement for illustrating the measurement technique.

The circuit of Fig. 1 is a typical example of a full-wave high-speed magnetic servo amplifier with separately excited 2-phase motor. This single-stage arrangement, a self-balancing magnetic amplifier,[3,4] employs special combination of internal positive magnetic feedback and negative electric feedback. In accordance with the fundamental principle of single-stage self-balancing magnetic amplifiers, self-saturation and additional positive external feedback produce an effectively infinite gain that is highly degenerated by negative feedback resistance-coupled into the control-circuit loop.

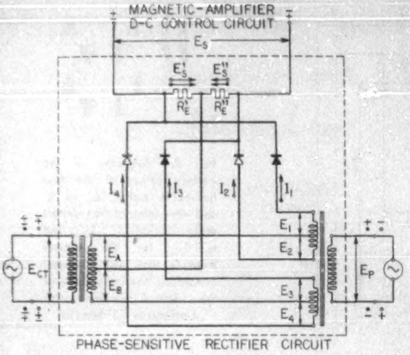
In the direct-voltage-controlled push-pull circuit of Fig. 1, the polarity-reversible signal voltage, E_S, is balanced by means of several voltage components. These components are supplied in time sequence by the polarity-reversible voltage drop, E_K, across feedback resistors R_K', R_K'' and by additional voltages induced in the control-winding elements from nonfired cores during their gating half-cycles. As a result of this fact, the balanced control circuit demands practically no energy from the d-c signal source, e.g., a phase-sensitive rectifier; and the effective input resistance is about 100 to 1,000 times higher than the actual input-circuit resistance.

In order to reduce size and weight of the amplifier, the comparatively large power-supply transformer used in similar arrangements[3,5,6] has been eliminated. This was made possible by employing a special form of a double-bridge type of self-saturating push-pull amplifier with two pairs of feedback resistors, R_K', R_K'' and R_E', R_E'', in the line circuit. In this case, the gate windings, N_A, of the four saturable-reactor elements are divided into two equally rated parts, which are insulated from each other.

The circuit arrangement of Fig. 1 has been operated from a 115-volt 400-cycle line voltage in connection with a standard-type 400-cycle 2-phase motor. This squirrel-cage motor has two separate 58-volt control windings W_L', W_L'' (maximum output power 3 watts or 20 volt-amperes) and the 115-volts line-field winding with series capacitor C_P securing proper phasing of auxiliary current I_P. Of course, it is also possible to replace the motor control windings W_L', W_L'' by two separate primary windings of an additional transformer and to connect the control winding of the motor across the secondary winding of this transformer. Application of an output transformer prevents the quiescent currents (load currents I_L' and I_L'' with no-signal conditions) from flowing through W_L' and W_L'' and corresponding heating of the motor. It also permits impedance-matching of the circuit components. It is to be noted, however, that the output transformer should be constructed of high-remanence material to preclude reaction of this transformer upon the saturable cores from stored-energy feedback.[6]

The 2-mil Orthonol[7] tape cores of the four saturable-reactor elements have the following dimensions: outer diameter, $1^3/_8$ inches; inner diameter, $1^1/_8$ inches; tape width 1/4 inch. Each core has four multilayer toroidal windings of enamel-insulated copper wire, no. *31* (Brown and Sharpe gauge): output windings, $N_A =$ 2 by 2,000 turns (2 by 50 ohms); bias windings, $N_B = 400$ turns (12 ohms); control windings, $N_C = 300$ turns (9 ohms); boosting windings, $N_D = 50$ turns (1.5 ohms).

Four center-tapped selenium-rectifier sets each having 2 by 4 cells with 1-inch square plates have been used, but germanium of silicon diodes may be used equally well. The resistance values of a-c bias resistors R_B, feedback resistors R_K', R_K'' and R_E', R_E'' are not critical. Thus



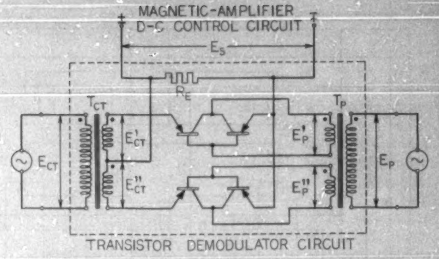


Fig. 3 (left). Star-modulator type of phase-sensitive rectifier circuit

Fig. 4 (above). Transistor demodulator circuit with auxiliary resistor, R_E

Both circuits are for the application of a polarity-reversible d-c signal voltage, E_S, to the input of the voltage-controlled self-balancing magnetic servo amplifier of Fig. 1

wire-wound or composition ("metallized") resistors may be employed in this circuit.

The graphical symbols of Fig. 1, which can be used in a general way for illustrating actual mode of operation of magnetic-amplifier circuits,[8] show that the black rectifiers are conducting simultaneously during the first half-cycle period of power-supply voltage E_P (black polarities), while the white rectifiers do conduct simultaneously during the second half-cycle period of E_P (white polarities). In this arrangement, feedback resistors, R_K', R_K'' and R_E', R_E'' in the line circuit, and motor control winding W_L', W_L'' are supplied from eight half-wave rectifier elements with variable unidirectional full-wave currents I_K', I_K'', I_E', I_E'', and I_L', I_L'' in such a way that the resultant d-c and a-c components of the four half-cycle pulses, I_1 to I_4, are polarity-reversible and are zero at zero input. Evidently, R_K' and R_K'' will carry unidirectional currents having the average values $I_K'=I_1-I_4$ and $I_K''=I_2-I_3$, whereas the motor control windings W_L', W_L'' are supplied with unidirectional currents $I_L'=I_1+I_4$ and $I_L''=I_2+I_3$, as already described in previous publications.[9-11] A common "a-c line resistor" R_T may be provided in this arrangement to limit actual magnitude of total alternating current I_T and to improve the operating conditions by reducing actual reverse voltage of the rectifier elements during the reset periods of the associated saturable-reactor elements.

In accordance with the previous graphical presentation,[11] the schematic diagrams of Fig. 2 illustrate the mode of operation of the circuit of Fig. 1 with regard to the 2-phase motor. It shows the variations of the average values of the four half-cycle pulses, I_1 to I_4, and the waveforms of the resultant motor-winding load currents I_L', I_L'', and $I_L'-I_L''$ for the typical example that, with no-signal conditions ($E_S=0$), the pulses I_1 to I_4 have 50 per cent of their maximum value, i.e., the firing angle of the four saturable-reactor elements is 90 degrees.

Evidently, with $E_S=0$, the average values of I_1 and I_4, and also those of I_2 and I_3 are equal, the fundamental-frequency components of I_L' and I_L'' are zero, and $I_L'-I_L''=0$. When a positive signal voltage, $+E_S$, is applied, the pulses I_1 and I_2 of the two "plus-minus saturable-reactor elements" will increase, while the pulses I_3 and I_4 of the two "minus-plus saturable-reactor elements" will decrease; and there is a resultant positive (0 degree) fundamental-frequency load current $I_L'-I_L''$ which is a function of the magnitude of $+E_S$. However, with negative values, $-E_S$, the pulses I_1, I_2 will decrease, the pulses I_3, I_4 will increase, and there is a resultant negative (180 degrees) fundamental-frequency load current $I_L'-I_L''$, which is a function of the magnitude of $-E_S$. Thus, the push-pull circuit of Fig. 1 will operate the separately excited 2-phase motor in the usual way by means of pure, phase-reversible alternating current $I_L'-I_L''$ (substantially without d-c component); and the feedback resistors R_K', R_K'' will produce the unidirectional, full-wave polarity-reversible negative-feedback voltage $E_K=I_K'R_K'+I_K''R_K''$ in the control-circuit loop of the amplifier. In this case, with $R_K'=R_K''=R_K$, the term $1/R_K=(I_K'+I_K'')/E_K=(I_K'+I_K'')/E_S$ represents the transconductance of the circuit with regard to the negative-feedback resistors.

Fig. 3 shows the circuit diagram of a star-modulator[12] full-wave type of phase-sensitive rectifier for converting the phase-reversible a-c error-signal voltage, $\pm E_{CT}$, of the synchro-control transformer into a polarity-reversible d-c signal voltage, $\pm E_S$. This arrangement contains the following:

1. The reference-voltage transformer T_P with ordinary core material and two center-tapped secondary windings.
2. The signal-voltage transformer T_{CT} with Supermalloy tape core and one center-tapped secondary winding.
3. Four silicon diodes.
4. Two mixing resistors R_E', R_E''.

If the reference voltages $E_1=E_2=E_3=E_4=\text{const}\times E_P$ are either in phase or 180 degrees out of phase with the signal voltages $E_A=E_B=\text{const}\times E_{CT}$ and if these reference voltages are larger than the maximum value of the signal voltages, then the polarity-reversible d-c signal voltage, $\pm E_S$, applied to the input of the test circuit of Fig. 1, will be proportional to the phase-reversible a-c error-signal voltage, $\pm E_{CT}$ of the synchro-control transformer.

In order to minimize zero-drift errors caused by large changes in power-supply voltage and ambient temperature, a transistor demodulator-circuit[13,14] may be used. In this device, Fig. 4, four p-n-p transistors having a Zener voltage of 42 volts are used as switches.[15] This type of demodulator is capable of operating at signal-voltage levels at least as low as 20 millivolts and as high as 30 volts rms with low drift over the temperature range from -65 to $+75$ degrees centigrade.[13]

When employing this transistor demodulator in connection with the d-c voltage-controlled amplifier of Fig. 1, a single auxiliary resistor R_E (Fig. 4) is provided to apply a polarity-reversible d-c control voltage, $\pm E_S$, to the input

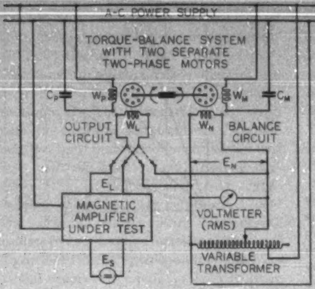


Fig. 5. Torque-balance system with load motor W_PW_L and balancing motor W_MW_N for measurement of output-input characteristics in terms of the equivalent sinusoidal output voltage E_N

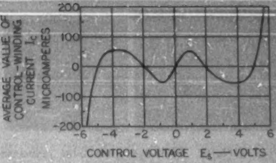


Fig. 7. Average value of control-winding current, I_C, as a function of d-c signal voltage, E_S, under critical-regeneration conditions

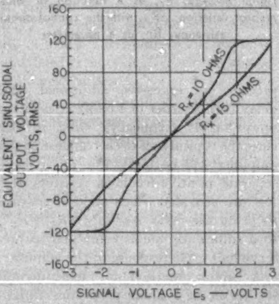


Fig. 6. Equivalent sinusoidal output voltage, E_N, as a function of d-c signal voltage, E_S, with feedback resistance, R_K, as a parameter

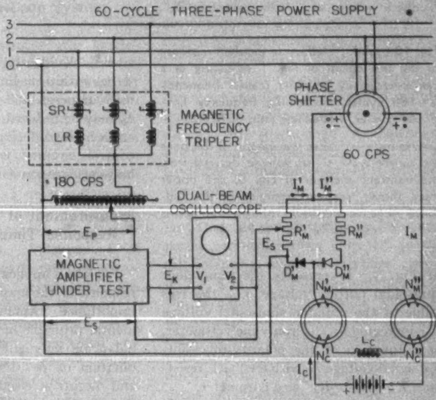


Fig. 8. Test circuit for measuring the transient - response time of high-speed magnetic servo amplifiers by means of a magnetic switch

of the amplifier. This is in contrast to the usual method, where the full polarity-reversible direct output current of the transistor demodulator flows through the control windings of the amplifier operating as a direct-current-controlled device.

Measurement of Output-Input Characteristics

Examination of the magnetic-amplifier circuit of Fig. 1 with regard to the possibility of measuring output-input characteristics makes it evident that conventional methods for making such measurements are not applicable on this arrangement and will also not be suitable in similar cases.

In order to make correct measurement of output-input characteristics on such arrangements possible, the torque-balance method was developed. It is based upon comparison of actual magnetic-amplifier output voltage with a continuously adjustable sinusoidal auxiliary voltage which has fixed phase relationship with respect to the power-supply voltage, preferably corresponding to maximum-torque operating conditions.

In accordance with the basic principle of the torque-balance method, the rotor shaft of the load motor W_PW_L has been flexibly coupled with that of an identical balancing motor W_MW_N, as shown in Fig. 5. Actual test procedure is as follows:

1. With the double-pole double-throw switch in dotted-line position, the variable substantially sinusoidal auxiliary voltage E_N is applied to the parallel-connected control windings W_L, W_N; and the variable capacitor C_M can be adjusted so that actual torque of motor system W_PW_L will be balanced by that of the opposing motor system W_MW_N. Since both torques are proportional to E_N, they will balance each other, when E_N is changed from zero up to its maximum value.

2. With the switch in solid-line position, actual highly nonsinusoidal output voltage E_L of the magnetic-amplifier output circuit is applied to control winding W_L of the load motor, whereas the variable sinusoidal voltage E_N of the balance circuit is applied to control winding W_N of the balancing motor. In this case, the torques of the two motor systems will balance each other, if E_L and E_N have equivalent effects.

3. Measurement of output-input characteristics can be achieved by varying signal voltage E_S (from zero up to its maximum value), and adjusting the variable transformer, i.e., voltage E_N so that the actual torque of system W_PW_L is balanced by the torque of system W_MW_N.

4. Graphical presentation of output-input characteristics consists then in plotting the equivalent sinusoidal output voltage E_N against signal voltage E_S, preferably with various circuit parameters.

Evidently, when following this test procedure, the torque-balance method offers the possibility of determining that value of sinusoidal voltage which produces on the 2-phase motor exactly the same torque as the actual highly nonsinusoidal output of the magnetic amplifier under test. This method which permits direct measurement of magnetic-amplifier output in terms of the equivalent sinusoidal output voltage has proved to be very practical for investigations on various full-wave and half-wave types of magnetic servo amplifiers.

Fig. 6 shows as an example a typical output-input characteristic of the circuit

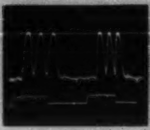


Fig. 9 (left). Half-cycle transient response of the d-c-controlled magnetic servo amplifier of Fig. 1

Lower trace indicates actual waveshape of the synchronous unidirectional square-wave signal voltage, E_S, derived from the transient analyzer of Fig. 8, and upper trace shows transient response of the full-wave negative-feedback voltage, E_K, across resistors R_K', R_K'' of Fig. 1

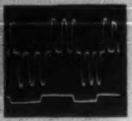


Fig. 10 (right). Half-cycle transient response of the test circuit of Fig. 1 operating as a carrier-frequency amplifier (carrier frequency f_c = 180 cycles, modulating frequency f_m = 60 cycles, frequency ratio $f_c/f_m = 3$)

Lower trace indicates waveshape of the pure a-c square-wave signal voltage derived from the transient analyzer of Fig. 8, and upper trace shows transient response of the full-wave negative-feedback voltage, E_K, across resistors R_K', R_K'' of Fig. 1

of Fig. 1. It presents the equivalent sinusoidal output voltage, i.e., the rms value of the variable sinusoidal voltage E_N, indicated by the iron-vane voltmeter, as a function of the polarity-reversible d-c signal voltage E_S with feedback resistance $R_K = R_K' = R_K''$ as a parameter.

The effective input resistance of the control circuit is dependent on the actual magnitude of control-winding current I_C (Fig. 1). Therefore, application of Supermalloy tape cores may be preferred in some cases to minimize actual magnetizing-current requirements of the control circuit. Since this amplifier does not contain any rectifier components in its control circuit, the control magnetizing current is capable of assuming any magnitude and direction necessary to support the actual magnetic-flux conditions imposed by the applied voltages. As a result of this very important fact, the amplifier will maintain its unique features even when employing substantially nonremanent core materials; and even lower-grade core materials may successfully be used in some cases.

Average value of control-winding current I_C has been measured by means of a shunted mirror-type galvanometer. Fig. 7 shows I_C as a function of polarity-reversible d-c signal voltage, $\pm E_S$, when a properly rated regeneration-control resistor R_D is provided in the differential-boosting circuit N_D', N_D''; R_D', R_E'', R_R (using "critical regeneration" conditions[3]) Evidently, average value of I_C varies within the limits of from 0 to ±100 microamperes, and the a-c component of I_C may be reduced to values of the same order by introducing a choke coil into the control circuit[16] without affecting the half-cycle response of the amplifier.

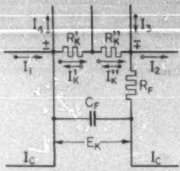


Fig. 11. Application of frequency-sensitive components (integrating circuit C_FR_F) in the degeneration network of the magnetic servo amplifier of Fig. 1, which permits elimination of the tachometer generator, as conventionally used for stabilization of closed-loop servo systems

Measurement of the Transient-Response Time

For servo applications, a high speed of response of the amplifier is of paramount importance. Application of the fundamental principle of the single-stage self-balancing magnetic amplifier, i.e., combination of positive magnetic feedback and negative electric feedback,[3] offers the possibility of reducing actual response time of the amplifier to its minimum value of 1 half-cycle of the power-supply frequency by introducing 100-per-cent negative feedback[16,17] without employing any rectifier components in the control circuit of the amplifier.

Analysis of the self-balancing magnetic-amplifier circuit of Fig. 1 reveals that half-cycle response will be obtained if turns ratio N_C/N_A (total) is approximately equal to the resistance ratio R_K/R_O, where $R_K = R_K' = R_K''$ is the actual resistance of the negative-feedback resistors, and R_O is the total resistance of each output-circuit component including the load-resistance elements, copper resistance of the gate windings N_A, and forward resistance of the rectifier units.

Referring to the use of the regeneration-control circuit, it is to be noted that this circuit, when improperly rated, i.e., if the ratio N_D^2/R_D is too large, will produce an appreciable additional time delay in the transient response of the amplifier. If N_D^2/R_D is too large, then this circuit—in a similar way as the control-circuit ratio N_c^2/R_c of a conventional full-wave magnetic-amplifier circuit—will produce an excessive additional time delay (in the order of 1 or 2 cycles of the power-supply frequency); and, in this case, the typical half-cycle response of this amplifier will be lost. For this reason, separate positive-feedback resistors R_E', R_E'' are provided which make it possible to bring the ratio $N_D^2/(R_D + R_E' + R_E'')$ to a sufficiently small value and to preserve half-cycle response of the amplifier.

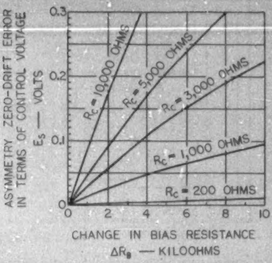


Fig. 12. Measured values of the asymmetry zero-drift error expressed in terms of d-c signal voltage, E_B, as a function of bias-resistor variation, ΔR_B, with the control-circuit resistance, R_C, as a parameter

A simple method for experimental studies on the dynamic properties of high-speed magnetic amplifiers consists in applying a synchronous square-wave signal voltage E_S derived from a motor-operated transient analyzer.[1] This device, which is operated by means of a mechanical switch, does permit direct visual observation of a stationary pattern. However, as already pointed out, application of this device is limited to low power-supply frequencies where a mechanical switch can be operated in a reliable way. For this reason, a novel type of synchronously operated transient analyzer was developed which may be used equally well in the range of higher power-supply frequencies, for example, those derived from magnetic frequency multipliers.[2]

The transient response of the circuit of Fig. 1 has been studied by connecting the amplifier with the 180-cycle output of a 60/180-cycle 3-phase type of magnetic frequency tripler (consisting of three saturable reactors SR, and three linear reactors LR).[2] The input of this tripler is supplied from the 60-cycle power supply which also feeds a phase shifter and a 2-core saturable-reactor circuit acting as a magnetic switch (Fig. 8).

The fundamental principle of this arrangement is based upon the fact that the actual wave shape of the alternating (load) current I_M of an ordinary saturable reactor with rectangular-hysteresis-loop core material (a transductor having series-aiding-connected a-c load windings N_M', N_M'' and series-opposing-connected d-c control windings N_C', N_C'') is nearly rectangular, if the control-circuit loop carrying constant direct current I_C represents a very high impedance (choke coil L_C) with regard to even-harmonic currents.[18–20] In this case, these even-harmonic currents are suppressed, and the saturable-reactor circuit does operate under forced-magnetization conditions.[21] In order to derive corresponding unidirectional square-wave pulses (signal voltage E_S) from the square-wave alternating current I_M, a splitting circuit[22] is provided. It consists of two half-wave rectifiers D_M', D_M'', with potentiometer resistor R_M' ("black" first-half-cycle pulses I_M') and fixed resistor R_M'' ("white" second-half-cycle pulses I_M'').

The oscillogram of Fig. 9 shows the transient response of the circuit of Fig. 1. It has been obtained by applying the synchronous square-wave signal voltage, E_S, derived from the magnetic-switch circuit of Fig. 8. The lower trace of Fig. 9 indicates this signal voltage, and the upper trace shows the transient response of the full-wave negative-feedback voltage E_K across resistors R_K', R_K''. Evidently, the response time is 1 half-cycle of the power-supply frequency, i.e., with $f_P=$ 400 cycles, the amplifier does reach 100 per cent of its final steady-state output current 1.25 milliseconds after the transient signal voltage E_S is applied; and the output current drops to zero 1.25 milliseconds after E_S is suddenly removed.

When employing the arrangement of Fig. 1 in connection with a 60/180-cycle or 400/1,200-cycle magnetic frequency tripler (without using a phase-sensitive rectifier in the control circuit) in a carrier-frequency system,[2] the line-field winding W_P of the motor will be separately excited in the usual way by a 60-cycle or 400-cycle current I_P. The motor control-winding components W_L', W_L'', however, will be introduced into the circuit of Fig. 1 in such a way that the pulses I_1, I_2 from the "plus-minus saturable-reactor elements" will flow through winding W_L', while the pulses I_3, I_4 from the "minus-plus saturable-reactor elements" flow through winding W_L''. In this case, the self-balancing half-cycle-response magnetic servo amplifier operates as a real a-c amplifier with a frequency-multiplier carrier frequency $f_c=180$ or 1,200 cycles a modulating error-signal frequency $f_m=$ 60 or 400 cps, and the frequency ratio $f_c/f_m=3$.

The oscillogram of Fig. 10 concerning this mode of operation has been obtained by applying the alternating voltage drop across the splitting circuit of the magnetic-switch arrangement, i.e., a pure alternating 60-cycle square-wave signal voltage, to the input of the circuit of Fig. 1. The lower trace of Fig. 10 indicates this signal voltage, and the upper trace shows the typical half-cycle transient response of the negative-feedback voltage, E_K, across resistors R_K', R_K'' corresponding to actual transient respose of the polarity-reversible output of the amplifier.

Measurement of Bandwidth on a 400-Cycle Closed-Loop System

Dynamic performance of the single-stage amplifier of Fig. 1 has been investigated by operating this amplifier in connection with a standard-type 400-cycle position servomechanism and the star-modulator[12] full-wave phase-sensitive rectifier of Fig. 3.

The half-cycle transient response of the arrangement of Fig. 1 does reduce the stabilization problems that usually accompany use of magnetic amplifiers in closed-loop servo systems. Measurement of bandwidth on a 400-cycle position servomechanism using a "servoscope" has shown that the bandwidth is about 17 cycles (107 radians), when a tachometer generator is employed in the usual way for stabilizing the servo system.[23] Bandwidth values of the same order have been obtained, when eliminating the tachometer generator by introducing frequency sensitive components[24] (e.g., the integrating circuit C_FR_F of Fig. 11) into the negative-feedback channel of the amplifier.

Measurements Concerning Reduction of Asymmetry Zero-Drift Errors

In order to illustrate the significant effect of drift-error reduction in self-balancing magnetic amplifiers, actual value of the asymmetry zero-drift error, expressed in terms of d-c signal voltage E_S has been measured: (1) without making any change in circuit configuration of Fig. 1; (2) providing the same voltage gain of the amplifier in all cases; (3) with various total-resistance values (R_C) of the control circuit; and (4) with various amounts of asymmetry. Well-defined amounts of asymmetry of the test circuit, Fig. 1, have been introduced by varying only one bias resistor (R_B) within the limits of from 30 to 40 kilohms, while the other three bias resistors (30 kilohms) remained unchanged.

Fig. 12 presents the measured values of asymmetry zero-drift error expressed in terms of d-c signal voltage E_S as a function of bias-resistor variation ΔR_B with the control-circuit resistance R_C as a parameter. The results of these measurements, Fig. 12, show in a dramatic way that, when reducing the total control-circuit resistance from the conventional value of 10,000 ohms to about 100 to 200 ohms and applying self-balancing operating conditions with critical regeneration (the effective input resistance E_S/I_C is at least 10,000 ohms), the actual magnitude of the asymmetry zero-drift error is reduced to about 1/40 of its original value. This is, of course, also true when the asymmetry of the push-pull circuit is caused by other effects, e.g., variations of the characteristics of the saturable-reactor and rectifier components. Thus, because the requirements on grading and matching of these components are greatly reduced, the single-stage self-balancing type of magnetic servo amplifier, Fig. 1, lends itself admirably to quantity-production techniques.

Finally, it is important to stress the fact that the drastic reduction of asymmetry zero-drift errors (Fig. 12) in the self-balancing magnetic-amplifier circuit of Fig. 1 is the result of combination of the dynamic gain-increasing effect of magnetic positive feedback and the drift-reducing effect of electric (galvanic) feedback, resistance-coupled into the control-circuit loop.[3,4] With regard to the great importance of this basic concept, H. F. Storm[26] has pointed out that "if one is confronted with the choice of either increasing the control-circuit resistance or applying negative electric feedback, it is well to remember that an increase of control-circuit resistance cannot change the drift, but an equivalent application of negative electric feedback will reduce the drift."

References

1. MAGNETIC-AMPLIFIER TRANSIENT ANALYZER, W. A. Geyger. *Electronics*, New York, N. Y., vol. 26, no. 7, Jul. 1953, pp. 180–91.

2. ELIMINATION OF ASYMMETRY ZERO-DRIFT ERRORS IN MAGNETIC SERVO AMPLIFIERS, W. A. Geyger. *AIEE Transactions*, vol. 74, pt. I, May 1955, pp. 230–45.

3. MAGNETIC AMPLIFIERS OF THE SELF-BALANCING POTENTIOMETER TYPE, W. A. Geyger. *Ibid.*, vol. 71, pt. I, 1952, pp. 383–95.

4. SELF-BALANCING MAGNETIC AMPLIFIER, W. A. Geyger. *United States Patent no. 2,700,130*, Jan. 18, 1955.

5. MAGNETIC AMPLIFIERS OF THE BALANCE DETECTOR TYPE—THEIR BASIC PRINCIPLES, CHARACTERISTICS, AND APPLICATIONS, W. A. Geyger. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1707–20.

6. FULL-WAVE REVERSIBLE-POLARITY HALF-CYCLE-RESPONSE MAGNETIC AMPLIFIERS, C. B. House. *Ibid.*, vol. 74, pt. I, Sept. 1955, pp. 541–52.

7. Magnetic Amplifiers with Orthonol Tape Cores, **W. A. Geyger.** *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 6, 1950, pp. 59–67.

8. Magnetic-Amplifier Circuits (book), **W. A. Geyger.** McGraw-Hill Book Company, Inc., New York, N. Y., 1954, chaps. 8, 9, 12–14.

9. A New Type of Magnetic Servo Amplifier, **W. A. Geyger.** *AIEE Transactions*, vol. 71, pt. I, Sept. 1952, pp. 272–80.

10. Two-Phase Induction-Motor Magnetic Amplifier with Direct-Current Braking, **W. A. Geyger.** *United States Patent no. 2,677,796*, May 4, 1954.

11. Reference 8, pp. 113–128, 132, Fig. 9.2 (c).

12. The Star-Modulator as a Twin-Push-Pull Type Modulator (Der Stern-Modulator als Doppel-Gegentakt-Modulator), **V. Aschoff.** *Telegraphen- Fernsprech- Funk- und Fernseh-Technik*, Berlin, Germany, vol. 27, no. 10, 1938, pp. 379–83.

13. Transistor Demodulator for High-Performance Magnetic Amplifiers in A-C Servo Applications, **R. O. Decker.** *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 121–23.

14. Transistor Switches Improve Magnetic-Amplifier Performance, **R. L. Bright, R. O. Decker, R. A. Ramey.** *Electrical Manufacturing*, New York, N. Y., no. 4, Apr. 1955, pp. 98–101.

15. The Use of Junction Transistors as Switches, **R. L. Bright.** *AIEE Transactions*, vol. 74, pt. 1, Mar. 1955, pp. 111–21.

16. Fast Response with Magnetic Amplifiers, **D. G. Scorgie.** *Ibid.*, vol. 72, pt. I, Jan. 1954, pp. 741–49.

17. On the Mechanics of Magnetic Amplifier Operation, **R. A. Ramey.** *Ibid.*, vol. 70, pt. II, 1951, pp. 1214–23.

18. A Simple D-C Type of Instrument Transformer Having Real Current-Transformer Properties (Ein einfacher Gleichstrom-Messwandler mit echten Stromwandler-Eigenschaften), **W. Krämer.** *Elektrotechnische Zeitschrift*, Berlin, Germany, vol. 58, 1937, pp. 1309–13.

19. Reference 8, p. 35, Fig. 3.4 (a).

20. Magnetic Amplifiers (book), **H. F. Storm.** John Wiley and Sons, Inc., New York, N. Y., 1955, p. 122, Fig. 8.4.

21. D-C Controlled Saturable Reactors and Their Feedback (Über gleichstromvormagnetisierte Wechselstrom-Drosselspulen und deren Rückkopplung), **T. Buchhold.** *Archiv für Elektrotechnik*, Berlin-Charlottenburg, Germany, vol. 36, 1942, pp. 221–38.

22. Reference 8, p. 114, Fig. 8.1 (b).

23. These 7 Steps Design a Tach Stabilized Servo, **J. E. Kadish.** *Control Engineering*, New York, N. Y., vol. 2, no. 8, Aug. 1955, pp. 72–79.

24. Theory of Servo Systems, with Particular Reference to Stabilization, **A. L. Whiteley.** *Journal*, Institution of Electrical Engineers, London, England, vol. 93, no. 34, pt. II, Aug. 1946, pp. 353–67. Discussion, pp. 368–72.

25. Magnetic-Switch Transient Analyzer, **W. A. Geyger.** *Electronics*, New York, N. Y., Jan. 1956, pp. 150–51.

26. Reference 20, pp. 213, 225.

27. The Operation of the Self-Balancing Magnetic Amplifier, **A. D. Krall, E. T. Hooper.** *AIEE Transactions*, vol. 75, pt. I, Mar. 1956, pp. 79–84.

High-Temperature Selenium Rectifiers—A Survey of Manufacturers' Data

T. S. SHILLIDAY
NONMEMBER AIEE

ABOUT a year ago the Subcommittee on Technical Information of the AIEE Metallic Rectifier Committee undertook a program for the purpose of accumulating information on high-temperature selenium rectifiers. As one part of that program Battelle Memorial Institute acting in behalf of the Metallic Rectifier Committee prepared and sent to selenium rectifier manufacturers a questionnaire on high-temperature selenium rectifiers. The 85-per-cent response to the questionnaire is an indication of the fine co-operation extended by the manufacturers. The information thus obtained was used to assemble a report to the sponsoring committee. That report, supplemented by additional information obtained through personal contacts with a few manufacturers, make up the contents of this paper. To avoid possible disclosure of proprietary information, the discussion will be kept on a general level. It is hoped, however, that a survey of the views, achievements, and direction of a major portion of the industry will be found of some value to both the manufacturer and the consumer.

Discussion of Rectifier Ratings

Each of the ten organizations from which a reply was received manufactures rectifiers for operation in ambient temperatures in excess of 50 C. Of these rectifiers, some are specially designed high-temperature cells and some are standard cells capable of high-temperature operation for limited periods with proper derating.

Data were submitted in a number of ways and in varying detail. Graphical means have been used to reduce some of these data to a common base. Relationships between expected life, rating, and ambient temperature are presented in the first four figures. Standard and high-temperature rectifiers have been treated separately to make clear the differences between them. In these figures, normal rating refers to a forward current density approximately equal to the National Electrical Manufacturers Association (NEMA) standard of 320 milliamperes per inch^2 for a single-phase bridge. Likewise, voltage per cell is, in most cases, 26 volts. Exceptions will be noted as they occur.

Fig. 1 shows current rating as a function of ambient temperature for standard rectifiers marketed by seven different manufacturers. Each symbol represents a different manufacturer. The table at the top of the graph lists the available information on expected life for each rectifier under various combinations of current rating and ambient temperature.

It will be noted that expected life at 50 C and lower ranges from 10,000 hours up for all rectifiers. At temperatures above 50 C, and for rectifiers falling within the area enclosed by dashed lines, reported life ranges from 1,000 to 3,000 hours. This area has been chosen rather arbitrarily to include the temperatures and ratings at which rectifier life is limited. With one exception, those rectifiers falling below the enclosed area have expected lives of 10,000 hours or greater. This increased life is presumably the result of more severe derating, and it is expected that lives of the order of 1,000 to 2,000 hours might be exhibited by these rectifiers at the higher ratings.

Fig. 2 shows voltage rating as percentage of normal rating at different ambient temperatures for the same rectifiers. A given manufacturer is represented by same symbol here as in Fig. 1. It can be seen here that voltage is not derated below 50 C and that usually, voltage derating is less severe than current derating.

Paper 56-77, recommended by the AIEE Metallic Rectifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing December 12, 1955.

T. S. Shilliday is with the Battelle Memorial Institute, Columbus, Ohio.

Grateful acknowledgment is made of the contributions to the subject matter of this paper by the following companies: Bradley Laboratories; Fansteel Metallurgical Corporation; Federal Telephone and Radio Company; General Electric Company; P. R. Mallory and Company; Radio Receptor Company; Sarkes Tarzian Inc.; Union Switch and Signal Company; Vickers Electric Division of Vickers, Inc.; and Westinghouse Electric Corporation. With one exception each of these organizations manufactures, in commercial quantities, selenium rectifiers for operation in ambient temperatures at 125 to 150 degrees centigrade (C).

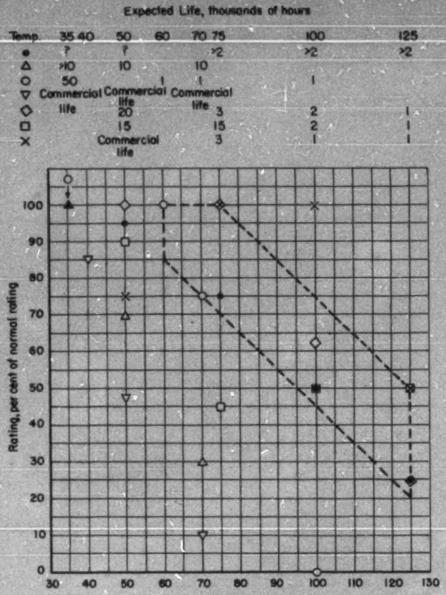


Fig. 1. Current derating and expected life for standard rectifiers at various ambient temperatures

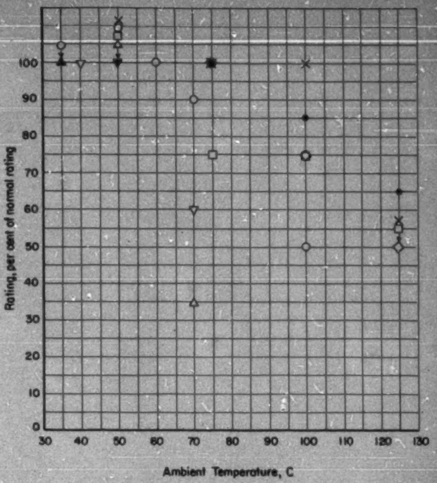


Fig. 2. Voltage derating for standard rectifiers at various ambient temperatures

It should be pointed out that the scatter in points on these two graphs and those to follow is not due entirely to product nonuniformity between manufacturers. This problem deals with four variables, current and voltage rating, ambient temperature, and expected life, none of which is independent of the others. With data coming from several sources it was inevitable that differences in manner of presentation would occur. Those data have been used here essentially as received. However, study of the graphs will show that some interpolation may be possible.

Of this group of seven manufacturers, those represented by the dot, the triangle, and the inverted triangle produce other selenium rectifiers specifically designed for operation at elevated temperatures. The manufacturers represented by the dot and inverted triangle reported that voltage ratings up to 40 volts per cell are available.

Fig. 3 shows the relationship between current rating, ambient temperature, and expected life for selenium rectifier cells designed for high-temperature operation. A given manufacturer is, in general, not represented by the same symbol in these high-temperature rectifier data as in the standard rectifier data.

In this graph the rectifier represented by the inverted triangle is a 24-volt cell, the one represented by the circle is a 45-volt cell, and the one represented by the square is available in 26- and 33-volt ratings. It will be noted that with two exceptions (the dot and the circle) the indicated life of high temperature rectifiers in ambient temperatures above 75 C is not significantly greater than that of standard rectifiers in the same temperature range. However, output current derating is generally much less severe and in two cases operation is extended to 150 C. The enclosed area from Fig. 1 has been superimposed on this graph to facilitate a comparison of the range of current ratings.

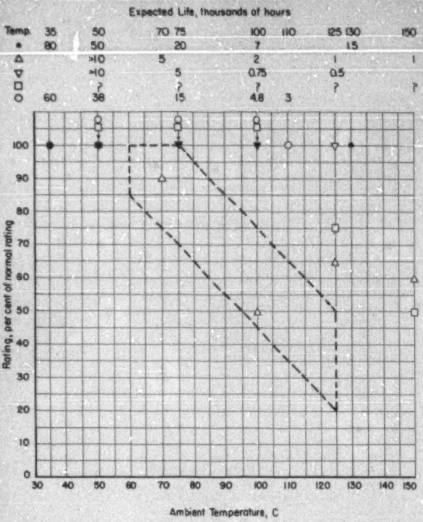
As shown in Fig. 4, only one of the manufacturers under discussion derates the input voltage of his high-temperature rectifier. This manufacturer has pointed out that his high-temperature rectifier is a rather new development and that until more extensive tests have been completed, the ratings on it are tentative and probably conservative.

An attempt was made to estimate rough averages of the current rating data available for standard and high-temperature rectifiers, taking into account expected cell life. The result is shown in Fig. 5. While these curves are not meaningful for the data of any given manufacturer, they do indicate the generally superior capabilities of rectifier cells designed for high-temperature operation.

Nearing completion in several laboratories are developments which should make available selenium rectifiers capable of several thousand hours operation at 125 C without derating. It is expected that plates such as these will operate without derating for 5,000 hours at 125 C. Life and rating at 150 C are as yet undetermined definitely but look very promising. One manufacturer reports a maximum ambient operating temperature somewhere between 150 and 200 C.

Factors Influencing Rectifier Aging

There was general agreement among manufacturers regarding the influence of output current density, input voltage, and ambient temperature on rectifier aging. An increase in any one of these factors accelerates aging; decreases in them reduce the rate of aging. Load current and cell temperature are the most important factors. Sufficient data are not available from the survey to define



Expected Life, thousands of hours

Temp.	35	50	70	75	100	110	125	130	150
●	80	50		20	7			1.5	
△		>10	5		2		1		1
▽		>10		5	0.75		0.5		
□		?		?	?		?		?
○	60	38		15	4.8	3			

Fig. 3. Current derating and expected life for high-temperature rectifiers at various ambient temperatures

clearly the extent of influence of input voltage on aging.

In discussing the relative effect of temperature and current density on rectifier aging one contributor reported that his test data suggested an approximately straight-line relationship between current density and forward aging rate in the range of cell temperature of 60 to 80 C. To the aging caused by current, this manufacturer adds the aging caused by cell temperature. It was pointed out that the effects of these two aging phenomena are approximately equal at 60 C. At 80-C cell temperature, the effect of temperature is greater than that of current density and at still higher temperatures the proportionate effect of temperature becomes still greater. Another manufacturer places this critical temperature at 100 C, at which cell temperature starts to become more important than current density for his rectifier.

A third contributor suggests that internal temperature rather than cell temperature (in the sense of plate surface temperature) may be a more meaningful parameter in the aging phenomenon. For a fixed cell or surface temperature, changes in current density alter the temperature gradient in the plate. Therefore, the same surface temperature at different current densities may not necessarily indicate the same internal temperature. Good aging characteristics have been obtained on cells operated at excessive current densities when clamped between large water-cooled blocks of copper, suggesting that internal temperature gradient is the critical factor.

It is believed that these two viewpoints are compatible since, if cell surface temperature is permitted to increase beyond the critical point at a given current density, the internal temperature gradient in the cell may be reduced or the rate of dissipation of thermal losses in the cell may be reduced.

Detailed comparison of the characteristics of standard and high-temperature cells at a given temperature were not obtained. However, three of those reporting stated that forward voltage drop

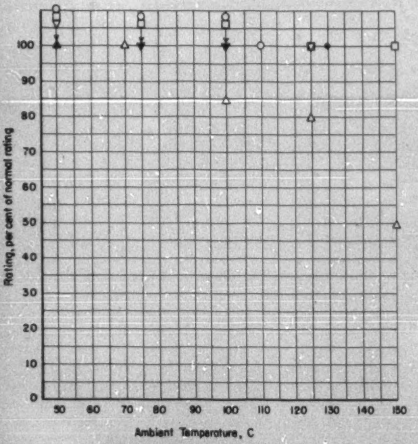


Fig. 4. Voltage derating for high-temperature rectifiers at various ambient temperatures

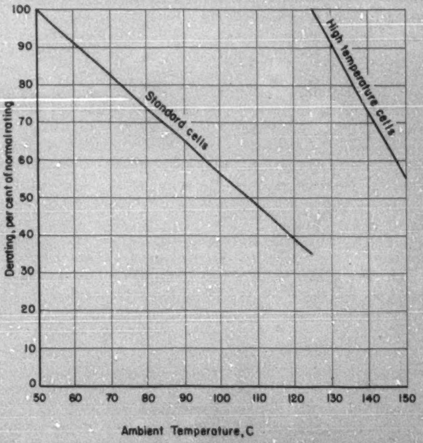


Fig. 5. Average current rating as a function of ambient temperature based on data available on standard and high-temperature selenium rectifiers

in their high-temperature cells at room temperature were 5, 10, and 60 per cent higher than their respective standard cells. This higher voltage drop, of course, decreases as ambient cell temperature is increased, because of the negative temperature coefficient of forward resistance.

Leakage currents of high-temperature cells in some cases change little or decrease slightly as ambient temperature increases. One manufacturer reported that compared to standard cells, his high-temperature cells exhibit very little unforming during storage. In these rectifiers the barrier layer is more stable and less affected by temperature and humidity.

Methods for Protection from Environment

Several methods are in use for protecting selenium rectifiers from contamination by ambient atmospheres. Depending on the application, industrial units are hermetically sealed, sealed in nylon enclosures, encapsulated in oil, embedded in a resin, or dipped or sprayed with a phenolic or vinyl-base enamel. The most widely reported protection from humidity or salt spray consists of a chlorinated rubber paint. One company uses a double coat of pigmented Bakelite varnish for protection from humidity, salt spray, or acid atmospheres. When protection from fungus is required, fungicidal varnishes are applied over other protective layers.

In evaluating the quality of a protective coating prior to its use in production the stated practices consist in general of:

1. Checking the compatibility of the finish and its solvent with the rectifier by an extended immersion test followed by electrical measurements on the rectifier. Where two types of paint must be used, their compatibility with each other is also checked.
2. Investigating the temperature characteristics of the protective coating by subjecting sample stacks to cyclic changes in ambient temperature, followed by examination of the electrical properties of the rectifier and the physical condition of the coating.
3. Studying the physical and electrical effect of humidity and salt spray on the rectifier and the coating.
4. Extensive life testing of rectifiers protected with the coatings passing the foregoing tests.

Summary

To summarize, four of the reporting manufacturers derate their standard rectifiers to permit operation at 125 C for periods up to 1,000 hours. Six manufacturers produce high-temperature rectifiers for operation at 125, 130, or 150 C. The expected life reported for the high-temperature cells is not significantly longer than that for standard cells at equivalent temperatures. However, high-temperature types do operate at much higher current and voltage ratings at elevated temperatures. It is also possible to rate cells at higher than NEMA ratings when used in short stacks or with wide spacing. Ratings exceeding the NEMA standard are also achieved by accepting reduced life expectancy, lower ambient operating temperature, or both.

The expected rectifier life figures cited in this paper are based on a 75- or 100-per-cent increase in forward voltage drop or a 10-per-cent decrease in output voltage. In many actual service applications, a much greater degree of aging may be permissible with correspondingly longer life. It is believed therefore that the ratings and life presented in this paper should be considered conservative.

Throughout the industry, research and development are being directed to many facets of selenium rectifier improvement. The high-temperature selenium rectifier is a major subject of these investigations. It appears at present that rectifiers capable of several thousand hours' operation without derating in a 125 C ambient will be available in the near future. Operation has already been extended to 150 C but much further improvement in performance at this temperature may be expected in the years to come. Because high-temperature cells are being designed for operation without derating at elevated temperatures, improved performance at normal temperatures should also be realized. Examples of this improvement are observed in Fig. 3 where 80,000 and 60,000 hours of useful life are reported for 26- and 45-volt high-temperature cells respectively.

◆

Discussion

J. Gramels (Bell Telephone Laboratories, Inc., New York, N. Y.): The author's survey on life expectancies and derating factors for high-temperature operation of selenium rectifiers is of great interest to the design and application engineer of power supplies. The data are a valuable guide to the engineer who is responsible for equipment that must meet a specified minimum number of hours of operation, particularly in military projects where higher ambient temperatures in the equipment are encountered because of confined space limitations.

It is also evident that there are considerable differences in manufacturers' derating factors and expected life. In this connection it may be interesting to report some studies made at the Bell Telephone Laboratories on selenium stacks operated at an 80 C ambient temperature.

This temperature was selected because it was somewhat in excess of the conventionally accepted theory that if selenium rectifiers were operated at cell temperatures above 75 C, a drastic reduction in life may be anticipated. The tests were conducted on the standard 26-volt type of selenium cells and are not to be confused with the more modern type of "high-temperature" cells.

The cells were operated at full-rated load current (320 milliamperes per square inch, single-phase full-wave circuit rating, NEMA standards). The a-c input voltage, however, was derated to 80 per cent of the maximum rating. Periodically, tests were made of the increase in the forward voltage drop at the rated current. As shown in Fig. 6, the data were plotted as per-cent increase from the initial forward drop. A 100-per-cent increase in this drop was chosen at the end of life (NEMA standards specify 75 per cent).

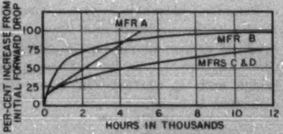


Fig. 6. Selenium rectifiers operated at +80 C ambient. Cells rated at 26 volts. Constant input voltage, 80-per-cent normal rating. Resistance load, 100-per-cent load current rating. Forward aging (a-c short-circuit test at rated current). Standard cells not processed for high-temperature operation. Data taken at room temperature

These results, as pointed out by the author, indicate a wide variation of life expectancy on different manufacturers' products. It also shows the difference in the rates of aging. The aging rate on stacks made by manufacturer *A* is essentially a linear function, indicating a life of about 5,000 hours. Stacks from manufacturers *B*, *C* and *D* exhibit an accelerated aging rate for the first few thousand hours, and then the aging rate tapers off for the duration of the test. The stacks from manufacturer *B* show an expected life of 12,000 hours, while stacks from manufacturers *C* and *D* have a life expectancy greater than 12,000 hours.

Cell temperature measurements under 80 C ambient temperature operation were not taken, but it is estimated that these temperatures were approximately 100 C.

Considering that: 1. These data were obtained on standard-type cells, not on high-temperature cells, and 2. The stacks were operated without current derating factors usually recommended at elevated temperature operation, the life expectancy obtained in these studies exceeded our expectations.

T. S. Shilliday: The information on which the paper is based consists, in general, of manufacturers' published rating data. Such data are believed by the author to be on the conservative side in most instances. Unfortunately, circumstances did not permit the performance of life tests at Battelle on representative samples of each manufacturer's rectifier. The value of such tests, however, in comparing the ultimate performance of rectifiers from various sources is clearly indicated by the curves presented by Mr. Gramels.

The "standard" rectifiers of the type upon which Mr. Gramel's data are based are covered in Figs. 1 and 2. A comparison of these data will show that the life of manufacturer *A*'s rectifier is reasonably typical of that which might be expected from the rectifiers represented by the diamond and the cross under similar operating conditions. However, the performance of the rectifiers of manufacturers *B*, *C*, and *D* far surpasses that predicted by data used in the survey.

The results of this comparison corroborate the opinion that published rating data are generally conservative, and point up the value of test data to supplement rating information for specified conditions, providing the test data are obtained from representative samplings.

The Use of an A-C Bridge to Measure Core Loss at High Inductions

I. L. COOTER
MEMBER AIEE

W. P. HARRIS
NONMEMBER AIEE

A KNOWLEDGE of the total core loss and its components is of vital importance to design engineers who use ferromagnetic materials which are subject to magnetic flux changes in direction or magnitude. Generally this information is obtained from measurements made on strip samples in an Epstein frame, with the use of a wattmeter. In many cases, however, there would be certain advantages, such as increased sensitivity and greater frequency range, in making these loss measurements by a bridge method. For measurements at low inductions and high frequencies the bridge methods have been successfully used for several years.[1,2] However, if the induction is increased over 10 to 12 kilogausses in nonoriented silicon steels or 16 to 18 kilogausses in oriented silicon steels, the distortion in the wave form of the exciting current is considerable, and it has been found that the results obtained with the bridge methods hitherto used differ from those obtained with the wattmeter. This difference increases rapidly with increases in induction. The bridge method always indicates larger power losses.

The discrepancies between results obtained at high induction by the two methods have generally been recognized as being caused by harmonic components in the exciting currents which result from the nonlinear element (iron-cored coil) in the circuit. Peterson[3] analyzed the problem of measuring the impedance of a nonlinear circuit element and suggested an approach to the solution of the problem. He assumed that the nonlinear element was equivalent to passive elements of resistance and inductance and a series of generators of harmonic frequencies. The element absorbs energy at fundamental frequency and delivers energy to the circuit at harmonic frequencies. The energy dissipated in the element will be the difference. This approach has been experimentally investigated[4,5] but with only partial success. This paper describes a more comprehensive application of Peterson's harmonic generator concept to the bridge problem, and shows that excellent agreement may be obtained in core loss measurements between the bridge and wattmeter even when the distortion is high.

Wattmeter Method

The most generally accepted way of measuring total core loss in sheet material at power frequencies is by the wattmeter method as specified by the American Society for Testing Materials.[1] The 25-centimeter (cm) Epstein test requires flat-rolled material cut into strips 3 cm wide and 28 to 30 cm long. The strips are assembled in the four arms of the test frame with double-lap joints at the corners. The diagram of connections for the core loss test is shown in Fig. 1.

W is the wattmeter the current coil of which is connected in series with the primary winding of the test frame. The voltage coil is connected to the secondary winding of the test frame. The average-indicating voltmeter is used 1. for determining the maximum induction in the specimen in accordance with Camilli's method,[6] and 2. in conjunction with the rms-indicating voltmeter for calculating the form factor of the secondary voltage.

Bridge Method

The bridge method consists essentially in measuring the resistive component of the impedance of the coil containing the test specimen, and the exciting current. From these values and the ohmic resistance of the coil, the power loss is computed. If R_d is the apparent increase in the resistance of the coil due to the core loss in the specimen and I is the current in the winding, the power loss is taken to be I^2R_d. This simple calculation is valid, however, only if the effect of harmonic components in the exciting current is so small that it can be disregarded. This condition obtains only at very low inductions.

Several different types of bridge[1] have been used for measuring core loss. In the present investigation, however, attention was confined to the Maxwell-Wien circuit illustrated in Fig. 2.

R_a and R_c = bridge product arms
R_{dt} and L_d = measured values of resistance and inductance of the self inductor
R_b and C_b = balancing resistance and capacitance
R_s = resistance of the source

Paper 56-26, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 19, 1955; made available for printing November 10, 1955.

I. L. COOTER and W. P. HARRIS are with the National Bureau of Standards, Washington, D. C.

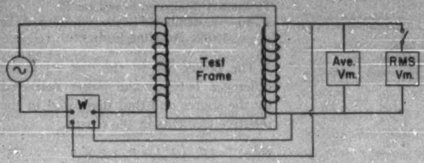


Fig. 1 (above). Connections for core loss test using wattmeter

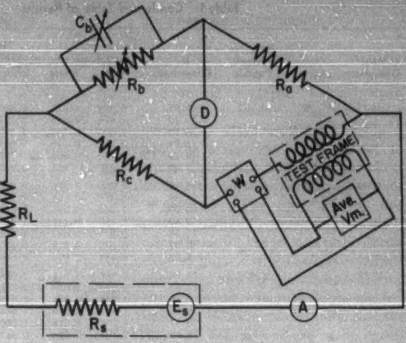


Fig. 3 (right). Comparison circuit for measuring core loss with wattmeter and bridge simultaneously

R_L = resistance of the leads, contacts, meters, etc.

D = tuned detector, sensitive only to the fundamental frequency in ordinary bridge measurements

The balance equations are

$$R_{dt} = \frac{R_a R_c}{R_b}$$

$$L_d = C_b R_a R_c$$

Simultaneous Measurements with Bridge and Wattmeyer

To compare directly the results obtained by the bridge with those obtained with a wattmeter, the circuit shown in Fig. 3 was used. A mutual inductance was used to compensate the air flux, but is not shown. A simplified diagram showing the various components involved is given in Fig. 4. It can be seen that the current coil of the wattmeter is simply inserted in the bridge arm containing the Epstein frame. Thus, by energizing the potential coil of the wattmeter from the secondary winding of the test frame, simultaneous measurements can be made by the two different methods. The wattmeter was a sensitive light-beam electrodynamic instrument which had been calibrated at power factors down to zero and frequencies up to 300 cycles per second. The results obtained with the wattmeter were taken as the standard values with which to compare the results obtained by the bridge.

The circuit is equivalent to Fig. 4, where

R_{wm} = resistance of the wattmeter

R_w = resistance of the primary windings of the frame and compensating mutual inductance

R_d = excess of the apparent resistance of the test winding over its ohmic resistance

R_{dt} = observed resistance

L_d = apparent inductance of that arm of the bridge

When the bridge is balanced

$$R_{dt} = \frac{R_a R_c}{R_b}$$

where $R_{dt} = R_{wm} + R_w + R_d$. From this, $R_d = R_{dt} - R_{wm} - R_w$.

If, in the bridge circuit shown in Fig. 4, the sum of R_a and R_b is made so much greater than the sum of the resistance of the other arms of the bridge that the power loss in the arms containing R_a and R_b is negligible with respect to the total this part of the circuit may be disregarded in the calculation.

The power delivered by the source is

$$P_s = \sum_{1}^{k} I_h^2 (R_d + R_w + R_{wm} + R_c + R_s) \qquad (1)$$

where $R_s = R_L + R_s$ and I_h = harmonic current in the circuit due to the electromotive force of the source E_h. Also

$$P_s = \sum_{1}^{k} E_h I_h \cos \theta_h \qquad (2)$$

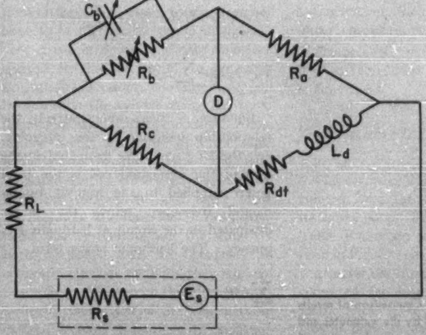


Fig. 2. Maxwell-Wien bridge circuit

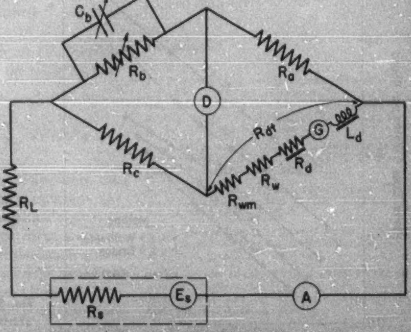


Fig. 4. Equivalent Maxwell-Wien bridge circuit containing wattmeter and Epstein test frame

Table I. Condensed Table of Results

AISI* Designation	Gauge	Effective Weight, Pounds	B_1, Kilogausses	$I_1{}^2R_d$, Watts	$\Sigma I_h{}^2R_p$, Watts	Power Loss, Watts per Pound: Bridge	Power Loss, Watts per Pound: Wattmeter	Δ, Per Cent
M-15	29	0.903	14.99	6.438	4.627	2.005	1.990	+0.8
			14.19	2.878	1.405	1.631	1.614	+1.1
			13.11	1.276	0.141	1.256	1.262	−0.5
			11.94	0.922	0.016	1.003	0.996	+0.7
			10.79	0.725	0.005	0.797	0.797	0.0
			9.64	0.587	0.002	0.648	0.642	+0.9
M-19	29	0.744	15.32	5.366	3.677	2.272	2.278	−0.3
			14.36	2.478	0.994	1.996	1.966	+1.5
			13.09	1.342	0.163	1.586	1.560	+1.7
			12.80	1.220	0.117	1.484	1.480	+0.3
			11.58	0.906	0.033	1.174	1.157	+1.5
			10.09	0.659	0.012	0.870	0.861	+1.0
			8.69	0.484	0.004	0.646	0.646	0.0
M-22	26	0.836	15.76	6.142	3.911	2.668	2.674	−0.2
			14.82	2.995	1.026	2.355	2.330	+1.1
			13.35	1.738	0.134	1.918	1.914	+0.2
			12.06	1.344	0.039	1.561	1.555	+0.4
			10.85	1.086	0.016	1.280	1.268	+0.9
			8.62	0.707	0.004	0.841	0.837	+0.5
M-36	26	0.877	15.74	8.022	3.913	4.684	4.720	−0.8
			14.59	4.153	1.089	3.493	3.511	−0.5
			13.88	3.027	0.418	2.974	2.964	+0.3
			13.03	2.323	0.123	2.508	2.508	0.0
			11.88	1.835	0.032	2.055	2.052	+0.1
			10.64	1.506	0.012	1.703	1.687	+0.9
			9.46	1.197	0.005	1.359	1.345	+1.0
			8.46	0.968	0.003	1.100	1.094	+0.5
M-43	24	0.895	16.41	10.254	3.912	7.084	7.020	+0.9
			15.42	6.125	0.960	5.769	5.725	+0.8
			14.15	4.251	0.174	4.554	4.468	+1.9
			12.85	3.242	0.036	3.581	3.574	+0.2
			11.52	2.537	0.009	2.824	2.815	+0.3
			10.29	2.007	0.004	2.237	2.234	+0.1

*American Iron and Steel Institute.

These data are not corrected for instrument loss or form factor.

where θ_h is the angle by which the harmonic component of current I_h lags the harmonic component of voltage E_h.

If the voltage generated by the source contains no harmonics of significant magnitude, only the voltage of fundamental frequency needs to be considered, and

$$P_s = I_1{}^2R_d + I_1{}^2R_p \tag{3}$$

where $R_p = R_w + R_{wm} + R_c + R_e$

Also

$$P_s = E_1I_1 \cos \theta_1 \tag{4}$$

Let

P_d = power loss in the core

P_{wm} = power dissipated in current coil of wattmeter

P_w = power dissipated in the winding of the test frame due to its ohmic resistance

P_c = power dissipated in resistor R_c

P_e = power dissipated in the resistance of the source including leads, etc.

The power loss in the iron P_d is the difference between the power delivered by the source and that dissipated in the resistive elements in the circuit.

$$P_d = P_s - (P_w + P_{wm} + P_c + P_e) \tag{5}$$

Although the voltage of the power source is known to be practically free of harmonics, the current may contain large harmonic components when the circuit contains a nonlinear element. Then

$$P_w + P_{wm} + P_c + P_e = \sum_1^k I_h{}^2(R_w + R_{wm} + R_c + R_e)$$

and from equation 4 $P_s = E_1I_1 \cos \theta_1$. The result is

$$P_d = E_1I_1 \cos \theta_1 - \sum_1^k I_h{}^2(R_w + R_{wm} + R_c + R_e) \tag{6}$$

Since only odd harmonics can be produced in the circuit by the iron-cored coil, then

$$\sum_1^k I_h = I_1 + I_3 + I_5 + I_7 + \ldots + I_k$$

$$\sum_1^k I_h{}^2(R_w + R_{wm} + R_c + R_e) = \sum_1^k I_h{}^2R_p = I_1{}^2R_p + \sum_3^k I_h{}^2R_p$$

Then

$$P_d = E_1I_1 \cos \theta_1 - I_1{}^2R_p - \sum_3^k I_h{}^2R_p$$

From equations 3 and 4

$$E_1I_1 \cos \theta_1 = I_1{}^2R_d + I_1{}^2R_p \tag{7}$$

$$P_d = I_1{}^2R_d + I_1{}^2R_p - I_1{}^2R_p - \sum_3^k I_h{}^2R_p$$

$$P_d = I_1{}^2R_d - \sum_3^k I_h{}^2R_p \tag{8}$$

Equation 8 shows quantitatively the relationship indicated in the foregoing, i.e., the true power P_d dissipated in the iron of the iron-cored coil is the total power supplied to the iron at fundamental frequency minus the power dissipated in the circuit at harmonic frequency. The harmonic power term

$$\sum_3^k I_h{}^2R_p$$

can be considered as being supplied by harmonic generators in the iron.

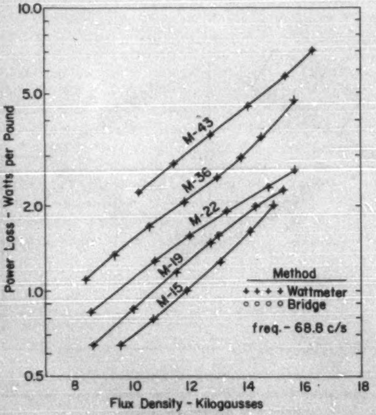


Fig. 5. Comparison of results obtained by the wattmeter and bridge for several grades of material

It should be noted that equation 8 is derived on the basis that R_p is independent of frequency. At power frequencies this assumption is permissible in so far as R_w, R_{wm}, R_e and R_L are concerned, but may not always be correct in the case of R_s. A more accurate equation is

$$P_d = I_1^2 R_d - \sum_3^k I_h^2 (R_w + R_{wm} + R_e + R_L) - \sum_3^k I_h^2 (R_s)_h \quad (9)$$

For the power source used in this investigation, however, R_s was found to be practically independent of frequency so that equation 8 will give valid results.

A wattmeter properly connected to the iron-cored coil will read the correct power dissipated in the core since it is an integrating device, and will automatically perform the operation of subtracting the power delivered by the coil at harmonic frequency, from the power supplied to the coil at fundamental frequency. However, any power measuring device which measures power at fundamental frequency only, will read the input power.

The power is supplied to the circuit by a source consisting of an oscillator having low distortion and a 200-watt power amplifier. This source was found to have a voltage wave form with a total distortion of less than 0.2 per cent (%) when delivering a current of 2 amperes to a resistive load. When connected to the bridge circuit, the apparent voltage distortion did not exceed 1.3% even though the current distortion was more than 50% in some cases. This increase in apparent distortion in the voltage wave is attributed to the $I_h R_s$ drop.

This circuit has several distinct advantages for the study of bridge measurements. The total resistance is low so that the voltage of the power source does not have to be high or the power dissipation very great in order to obtain sufficient current in the test frame to approach saturation induction. Measurements can be made on the sample using the wattmeter and bridge simultaneously. These measurements are made with the same flux density in the sample and with the same exciting current having identical harmonic components and wave form. Furthermore, the form factor and instrument loss corrections are the same for both methods and, for purposes of comparing the two methods, they need not be applied. The circuit therefore permits a direct comparison of the two methods under identical conditions of measurement and without requiring the application of the usual corrections.

Observations and Results

A 25-cm Epstein frame was used in the measurements of core loss for this investigation. It was built according to the specification of the American Society for Testing Materials with a total of 700 turns in the primary winding and 700 turns in the secondary winding. An inductor was used to compensate for air flux.

The specimens used in the frame were 3 cm wide and 28 cm long; the weight of each sample was approximately 500 grams. Five types of material were used with the following designations:

AISI* Designation	Grade	Approximate Silicon Content
M-15	transformer	3.5% to 5%
M-19	transformer	3.5% to 5%
M-22	dynamo	2% to 3.5%
M-36	electrical	0.5% to 2%
M-43	armature	0.5% to 2%

* American Iron and Steel Institute

Samples of these materials were inserted in the 25-cm Epstein frame using double-lapped joints. The primary or magnetizing winding of the test frame was connected in series with the primary winding of a compensating mutual inductor and the current coil of the electrodynamic wattmeter. Together these formed one arm of the Maxwell-Wien bridge for which the simplified diagram is shown in Fig. 4.

The fundamental and harmonic components of the current in the primary winding of the test frame are obtained from measurements of the voltage across the standard resistor R_c, made with a wave analyzer tuned to respond only to the frequency of the particular harmonic under consideration.

The general equation used for calculating total core losses by the bridge method is equation 8.

$$\text{Total core loss} = I_1^2 R_d - \sum_3^k I_h^2 R_p = P_d$$

The terms of this equation were calculated from the data obtained in the measurements. The results are given in Table I and are presented in graphical form in Fig. 5. Neither the data from the wattmeter nor from the bridge have been corrected for instrument loss or form factor since these results are for comparison purposes only. The final column of Table I and the curves show that the two methods are in excellent agreement.

Summary and Conclusion

The use of an a-c bridge for measuring core loss has some advantages over the conventional wattmeter method, such as a wider frequency range and better sensitivity. However, it has been found that the bridge method is subject to errors which may be large if the effect of distortion in the magnetizing current is neglected. The correction for this may be obtained by considering the equivalent circuit for the ferromagnetic circuit element to consist of resistance, inductance, and a series of generators delivering current to the circuit at harmonic frequencies. This method requires that the apparent resistance of the entire circuit connected to these generators be determined, including the apparent resistance of the power source. The power dissipated in the ferromagnetic material is the power supplied to the material by the source at fundamental frequency minus the harmonic power the material returns to the circuit.

A circuit was devised by which measurements of core loss can be made by the bridge and wattmeter simultaneously. By so doing, discrepancies due to differences in frequency or induction are obviated and comparisons can be made at identical values of frequency and induction.

Excellent agreement between the two methods was obtained for inductions as high as those for which the wattmeter is considered to be accurate.

It is therefore concluded that even for measurements at higher inductions than have hitherto been considered feasible, accuracy of an a-c bridge is comparable with that of the wattmeter, provided that the corrections for harmonic components in the exciting current are determined and applied with care.

References

1. Standard Methods of Test for Alternating Current. Core Loss and Permeability of Magnetic Materials. *Standard (A343-49)*, American Society for Testing Materials, Philadelphia, Pa., pt. 1, "Ferrous Metals," 1952.

2. A Wide Range A-C Bridge Test for Magnetic Materials, D. C. Dieterly, C. E. Ward. *Ibid., TP88. Bulletin 182*, May 1952.

3. Impedance of a Non-Linear Circuit Element. **Eugene Peterson.** *AIEE Transactions*, vol. 46, May, 1927, pp. 528–34.

4. Harmonic Power in Iron Testing. **J. Greig, J. E. Parton.** *Engineering*, London, England, July-Dec. 1938, p. 432.

5. Iron-Loss Measurements by A-C Bridge and Calorimeter, **J. Greig, H. Kayser.** *Journal*, Institution of Electrical Engineers, London, England, vol. 95.2, 1948, pp. 205–16.

6. A Flux Voltmeter for Magnetic Test, **G. Camilli.** *AIEE Transactions*, vol. 45, May, 1926, pp. 721–28.

Discussion follows

Discussion

D. C. Dieterly (Research Laboratories, Armco Steel Corporation, Middletown, Ohio): This is a fine piece of work which demonstrates so convincingly that wattmeter and bridge methods can give equivalent values for core loss when the tests are conducted under similar conditions. Those of us who make core loss measurements at high inductions by wattmeter methods for grading of electrical steel, and also bridge tests at certain other inductions, have been aware that bridge methods can give fractional per-cent agreement with wattmeter methods when the test induction is limited to flux densities where the harmonic power is negligibly small compared to that for the fundamental frequency. In bridges designed recently we have provided much smaller resistance in the arm carrying the specimen's exciting current than was done years ago, and have also taken pains to keep the resistance of the test winding and of the power source low. We find that harmonic power can be made small enough to be neglected up to 12 kilogausses on conventional grades of electrical steel, and up to 16 kilogausses on the oriented product. Although we are not inclined to push the bridge test method to higher inductions because of the difficulties which then become of paramount importance, the authors have shown that it can be done.

Hesitancy to employ bridge methods at moderate and reasonably high inductions seems to stem from the belief that a large value of resistance is a necessity in series with the specimen, and that the resulting distortion of the flux wave form necessarily makes the results substantially in error. Also, under such conditions the user often questions whether the desired value of B_{max} was being achieved in the test, and whether the use of a balance only on the fundamental frequency was justifiable.

The authors have wisely incorporated the basic features of the Epstein wattmeter method into the bridge test, thereby preventing some procedural differences from clouding the issue. With the sensitive detection devices now available, the resistance in series with the unknown need not be much larger than that of the test winding itself. It might be well to point out to those not familiar with the work of Camilli (see reference 6 of the paper) that the use of the average-sensitive voltmeter (fluxvoltmeter) on an essentially unloaded secondary winding ensures that the correct value of B_{max} is reached if the test induction is calculated from the indications of such a meter, and likewise ensures that the hysteresis component of losses will be essentially the same as that for sinusoidal flux wave shape. Use of the true rms-reading instrument in conjunction with the fluxvoltmeter to determine the form factor of the induced voltage enables one to calculate the error in eddy-current component of losses caused by distortion of the induced voltage. Although this was not done in this work, not being a factor of difference in the two results, it is required procedure for tests made by the wattmeter method in accordance with American Society for Testing Materials practice. As for the question of balance only at the fundamental frequency, one must remember that in wattmeter measurements with essentially sinusoidal voltage on the potential coil windings, only power at fundamental frequency will produce torque on the instrument. Therefore the only difference (and the only factor which makes for difficult handling by the bridge method as it is currently used) is that harmonic distortion in the series resistance arm is reflected in the balance parameters and must be evaluated when of significant magnitude, as it was in the work undertaken by the authors.

In general, we believe that the simplicity of wattmeter measurements will probably confine the practical use of bridge tests at high inductions almost entirely to cases where wattmeter methods lack sensitivity or frequency range. Although rather costly a good reflecting type of wattmeter of very high sensitivity is available for use at frequencies up to 15 kc which enable wattmeter methods to be used over a greater range than is generally recognized. Having had both types of equipment available in the same laboratory for a number of years, we are inclined to use the bridge method for all a-c testing where data at 10 kilogausses (15 kilogausses on oriented materials) or less are of prime interest, and the wattmeter method where 10 kilogausses and higher are the inductions of interest.

While we believe that the modified Hay bridge circuit[1] offers much simpler procedure for obtaining values of magnetic properties by treating the magnetic core as a parallel combination of resistance and inductive reactance, and the winding resistance as a series resistance element, it is understandable that the authors chose the Maxwell circuit as a means of simplifying the proof in this particular approach. Perhaps someone will eventually find a way of taking care of the harmonic power in the bridge method which is simpler than the rather laborious procedure the authors have used successfully. Meanwhile, users of bridge methods will appreciate the wider field of usefulness encouraged for that instrument.

REFERENCE

1. See reference 2 of the paper.

I. L. Cooter and **W. P. Harris:** Mr. Dieterly points out that the bridge method has generally been limited to use at inductions of 10 kilogausses or less. This, of course, is a direct result of the disagreement between the results obtained by the wattmeter and bridge in core loss measurement at higher induction. The method described in the paper should clarify the situation and make possible the use of the bridge at the highest induction. The paper also emphasizes the need for power sources having sine wave voltage output and low impedances, and the necessity for keeping the resistance in the loop containing the test frame and power source as low as possible. For the experiments reported, the resistance, including the source, ratio arm, wattmeter, leads, and test frames, was only 4.3 ohms. If this resistance is increased, as by using a 10-ohm ratio arm, the harmonic power correction increases at any given induction and becomes significant at a lower induction.

However, now that we have a method for determining the harmonic power correction, there is no longer any fundamental bar to the use of the bridge at any induction. The choice between wattmeter and bridge methods can now be based on convenience and suitability.

Vol.
Pt. 1
1956

Nontracking Organic Insulations

R. S. NORMAN
NONMEMBER AIEE

R. A. PFUNTNER
MEMBER AIEE

A. A. KESSEL
NONMEMBER AIEE

Synopsis: A novel principle has been discovered which retards or completely eliminates the formation of conducting decomposition products in organic materials exposed to electric discharges. The application of this principle to several organic insulation materials has greatly increased their ability to withstand electric arcing.

MANY new and improved organic electric insulations have become available over the past decade. Recent emphasis has been on the development of materials with more versatile physical characteristics and with improved high-temperature properties. Comparatively little work has been published on the compounding and synthesis of organic materials to improve a fundamental weakness, namely, their tendency to carbonize and form a conducting track on exposure to electric arcing conditions. Olyphant,[1,2] and Yarsley and Grant[3,4] have made significant contributions to measurement techniques and to the understanding of the basic mechanism of track initiation and propagation, but their work has been primarily concerned with the available insulating materials.

This paper reports the discovery of a novel principle which greatly increases the resistance of organic insulations to damage from electric discharges by reducing or completely eliminating the formation of conducting decomposition products. This is accomplished by building into the organic material an oxidation mechanism which prevents carbonization. Laboratory test methods and results on insulating materials exposed to arcing conditions are reported; surface creepage in contaminated atmospheres, arc-tracking, and internal ionization failure are discussed.

Internal Oxidation Mechanism

Previous workers have pointed out that the resistance of organic materials to damage from electric arcing is largely determined by their initial chemical composition and the nature of the gaseous and solid decomposition products. Olyphant[1] stresses the importance of the ambient gas and gaseous products of decomposition, Yarsley and Grant[3] emphasize the amount and nature of the decomposition products, Warren[5] cites the adverse effect of materials with carbon-rich phenyl groups, and Albert[6] states that the formation of conducting carbon is retarded in the case of vulcanized fibre by the large percentage of oxygen in its cellulosic structure. This suggests that, as in ordinary combustion where a free supply of oxygen is effective in lessening the quantity of solid carbonaceous deposits, an adequate source of oxygen within an organic insulating material might substantially reduce the formation of carbonized tracks under arcing conditions. However, the successful application of this principle to specific organic insulations has not previously been reported in the literature.

The nontracking materials discussed herein utilize a unique and basic formulation. The active component, which contributes to the oxidation mechanism, is a material which can be physically interspersed throughout the bulk of the basic organic material. This operation normally takes place during the compounding of the fillers, plasticizers, etc., with the base material. Under the influence of electric discharge in close proximity with the insulation material containing this interspersed ingredient, a reaction takes place which has unusual effectiveness in preventing the accumulation of carbonaceous deposits.

It is believed that the high arc temperatures resulting from the electric discharges break down the organic components of the insulating material and simultaneously trigger the oxidation reaction to form volatile hydrocarbons and gaseous oxides of carbon. Because this prevents the formation of free carbon, no conducting track is formed, and the only effect is a very slow general erosion, see Fig. 1.

These results are confirmed by two general classes of tracking tests.[3] In the first class of test the resistance to surface creepage damage by localized sparking is measured under contaminated conditions. This simulates outdoor use where nontracking elastomers and casting resins would have immediate application.

The second class of tracking test measures resistance to an arc drawn between two electrodes on the surface of a clean material.

Surface Creepage

Surface creepage in contaminated locations is a primary concern of electric apparatus designers. Under conditions of moisture, salt fog, or dust, leakage currents may result which produce arcing on the exposed surface of the insulation.[7-9] Under these conditions conventional organic insulating materials break down to form localized carbon paths.

Laboratory tracking tests, to be valid, must closely duplicate the expected service conditions and, in addition, must be capable of rapid and reproducible results. Several workers have described accelerated laboratory creepage tests on porcelain insulators coated with moist synthetic dusts containing a small amount of electrolyte and an insoluble binder.[9-11]

Various types of tracking tests as performed on organic insulating materials are discussed by Yarsley and Grant.[3] Their discussion includes tests in which scintillations are initiated by passing current through dust deposits between electrodes and tests in which current is passed through a drop of electrolyte between electrodes.

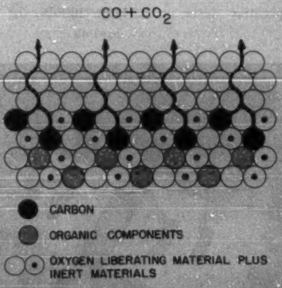


Fig. 1. Internal oxidation mechanism for elimination of tracking in organic insulating materials

---

Paper 55-699, recommended by the AIEE Dielectrics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 3-7, 1955. Manuscript submitted June 6, 1955; made available for printing September 16, 1955.

R. S. NORMAN, R. A. PFUNTNER, and A. A. KESSEL are with the General Electric Company, West Lynn, Mass.

Appreciation is expressed to M. W. Albright, D. E. Crawford, A. C. Bluestein, and the many others who contributed to this development. The encouragement and counsel offered by Dr. T. O. Paine during this investigation are greatly appreciated. Acknowledgement is also expressed to the Chemical Materials Department of the General Electric Company for compounding and evaluating phenolic materials utilizing the internal oxidation mechanism.

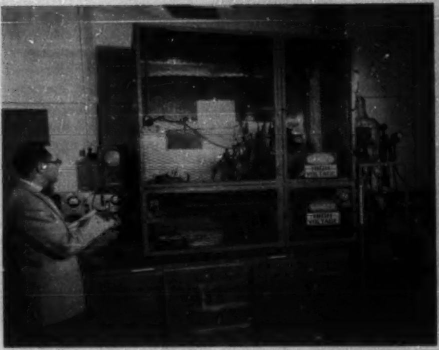


Fig. 2. Test equipment to study creepage resistance by dusted-sheet sample technique

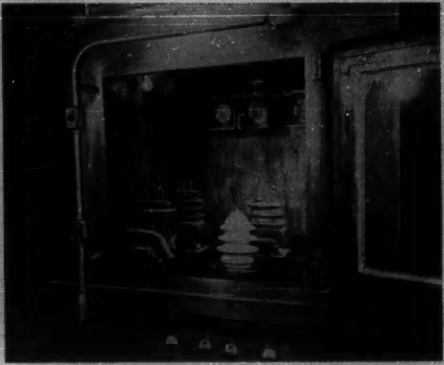


Fig. 4. Laboratory salt-fog chamber used primarily for creepage study of actual products

ACCELERATED SURFACE CREEPAGE TESTS

Based on this and other previous work, two modified tests were developed to evaluate creepage resistance under accelerated conditions. One test employs the moist dust principle and the other utilizes a salt fog.

In the dusted sample test, Fig. 2, a 75-mil sheet of the material to be tested is placed on a metal ground plate. Two parallel bar electrodes are spaced 1 inch apart on the insulation sheet. One electrode is connected to the ground plate, and a voltage of 1,500 volts is applied between the other electrode and ground. A partially conducting synthetic dust is placed on the insulation surface and continuously moistened with a tap-water spray. Under these conditions scintillations are constantly maintained between the electrodes; see Fig. 3. This relatively simple setup gives a rapid, reproducible means of making comparative tests on a large number of materials. However, the method is not readily applicable to actual apparatus having a complicated surface configuration.

The second accelerated test utilizes a salt-fog chamber continuously supplied with a dense fog of 10-per-cent NaCl and 1.2-per-cent MgCl. This chamber is used primarily for testing actual apparatus; see Fig. 4. A direct correlation between this test and severe coastal conditions is described in a companion paper.[7]

RESULTS OF ACCELERATED TESTS

Conventional Materials

Table I shows the time to failure of some typical organic insulations exposed to the dusted sample test. The results of tests on several samples of each material were very consistent except in the case of filled epoxide resin which is discussed in the following. The appearance of several samples is shown in Fig. 5.

It was observed that once an initial spot is carbonized from the arcing, this spot serves as a starting point for further degradation and a surface carbonized path forms at an ever-increasing rate. Because it is difficult to make filled epoxide samples without an appreciable resin skin on at least one surface, epoxide results tend to be erratic. When the tests are made on a surface having a heavy filler concentration (the bottom side as cast), the 40-hour time to failure is observed, but the other less filled surfaces invariably fail in less than 1 hour. The appearance of the melted glass in the melamine glass laminate sample, Fig. 5, indicates the high arcing temperatures produced by this test. In addition to many organic materials, silicone rubber was tested and gave excellent results as expected from its basic inorganic structure.

The surface tracking observed in the dusted sample test is also found in the salt-fog test. There is some variation between the resistance of materials to the two tests. For example, butyl

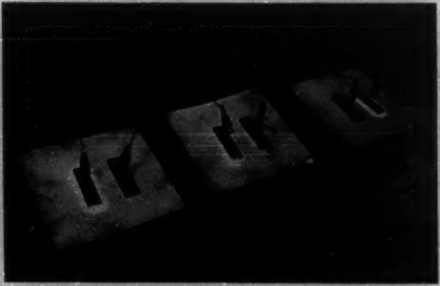


Fig. 3. Insulation sheet samples undergoing creepage test

Table I. Dust-Creepage Test on 75-Mil Conventional Organic Materials

Material	Hours to Failure
Silicone glass laminate	0.2
Filled polyester	0.2
Vulcanized fibre	0.3
Polyvinyl chloride	0.5
Neoprene	1.0
Melamine glass laminate	1.0
Hypalon *S-2*	1.0
Filled butyl rubber	2.0
Polyurethane	8.0
Melamine formaldehyde	12.0
Fluorinated rubber	18.0
Filled epoxide	0.5 to 40

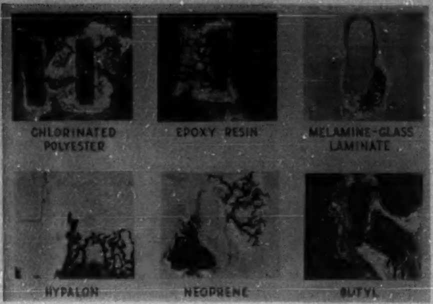


Fig. 5. Conventional organic insulations after exposure to the dusted sample creepage test

rubber is much more resistant to creepage damage in salt fog than epoxide resin. Transformers molded in conventional butyl rubber and cast in conventional epoxide resin are shown in Fig. 6 after exposure to accelerated arcing in the salt-fog test. Conventional butyl rubber transformers failed after 90 hours in the salt fog; the conventional epoxide transformers lasted less than 1 hour.

Based on the foregoing work, the following general sequence is suggested as leading to creepage failure of filled organic materials. The high localized temperatures produced by repeated sparking on the surface of the insulating material breaks down the organic components to form carbon. The exposed carbon at the surface is oxidized and goes off as a gas; see Fig. 7. The inert fillers left behind now form a blanket over the surface. As the remaining organic material breaks down, the blanket of filler prevents complete oxidation of the carbon. The unoxidized carbon accumulates and forms a conducting path under the inorganic filler; see Fig. 8.

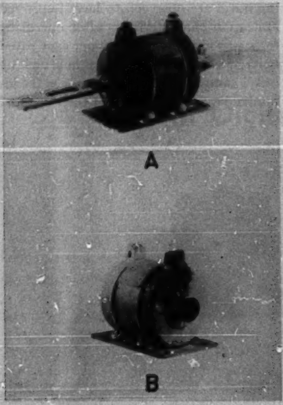


Fig. 6. Conventional butyl and epoxide insulated transformers after laboratory salt-fog test at 600 volts

A—Conventional butyl, surface-creepage failure in 90 hours

B—Conventional epoxide, surface-creepage failure in 1 hour

Fig. 9 shows an enlarged cross-sectional view of the failure path of a conventional butyl-insulated transformer after the salt-fog test. The appearance of the cross section provides evidence for the aforementioned mechanism. Although the protecting skin exists only with filled materials, the basic premise of carbon formation resulting from insufficient oxygen appears valid for all organic insulations.

Materials Utilizing Internal Oxidation Mechanism

Table II shows the improvement obtained on the dusted sample test by utilizing the internal oxidation mechanism in three conventional materials. The appearance of the samples after test is shown in Fig. 10.

No carbon spots form on butyl rubber or epoxide resin and, consequently, there is no concentration of the sparking along carbonized paths. Although a few carbonized spots were visible on the neoprene, in all cases failure ultimately occurred by erosion through the thickness of the samples. Obviously the time-to-erosion failure of the nontracking materials can be lengthened by increasing the thickness, while the time-to-tracking failure of conventional insulating materials is a function of the length of time required for a carbon track to progress across the surface of the insulation.

Similar results are obtained in the salt-fog tests. A typical example of the improvement obtained with the nontracking materials is illustrated by Fig. 11. One of these identical 600-volt transformers is insulated with conventional silica-filled epoxide resin, and the other with epoxide resin employing the internal oxidation mechanism. While the conventional epoxide insulation formed a carbonized path and failed after 1 hour at 600 volts in the salt-fog chamber, the nontracking material exhibited no carbon tracking, and finally failed by erosion after 1,400 hours under the same conditions. The life could have been extended even beyond this time by increasing the insulation thickness.

Fig. 12 shows a conventional butyl rubber transformer, a nontracking butyl rubber transformer, and a 15-kv porcelain bushing after exposure in the salt-fog chamber at 5 kv. Although the conventional butyl failed by carbon tracking in 107 hours, no carbonization was visible on the nontracking butyl after 1,400 hours. The failure of the 15-kv porcelain bushing, caused by cracking from thermal shock after 432 hours at 5 kv, emphasizes the severity of this accelerated test.[7]

A direct correlation between the accelerated salt-fog chamber tests and the most adverse service conditions found in the United States has been obtained for 5,000-volt butyl instrument transformers. Nontracking butyl does not carbonize

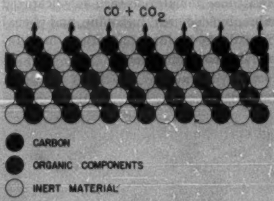


Fig. 7. Creepage failure of filled organic material, initial surface oxidation

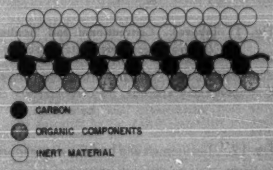


Fig. 8. Creepage failure of filled organic material, formation of carbonized path

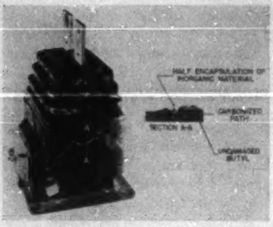


Fig. 9. Creepage failure of a conventional butyl transformer showing enlarged cross section of failure path

under either condition; conventional butyl, which fails in approximately 100 hours in the salt-fog chamber, lasts approximately 1,500 hours at the most severe outdoor location, an oceanside test station in San Francisco.[7]

Arc Tracking—ASTM

The elimination of carbonization by the internal oxidation mechanism is not only effective in reducing creepage damage but also greatly increases the arc resistance of organic materials. The standard test for arc resistance is the American Society for Testing Materials (ASTM) *495-48T* arc-resistance test,[12] in which a high-voltage low-current arc is drawn between two pointed electrodes on the surface of a clean material. This test is generally accepted for evaluating the arc resistance of materials such as vulcanized fibre, molded urea, melamine, and phenol formaldehyde for applications not involving contamination.

Table II. Dust-Creepage Test on 75-Mil Samples

Base Material	Hours to Failure	
	Conventional Filled	Utilizing Internal Oxidation Mechanism
Butyl rubber	2.0	400
Epoxide resin	0.5 to 40	400
Neoprene	1.0	31

The application of the internal oxidation mechanism to phenolics is of particular interest because of the widespread use of these materials. For example, W. Vassar and C. Leff[13] have recently drawn attention to the desirability of a low-cost arc-resistant phenolic-type material. Although phenolics are widely used by the electrical industry, their arc-tracking resistance is notoriously poor.[4] Previous workers have suggested that the high carbon content of the phenyl group[5] and the tarry decomposition products contribute to this poor tracking resistance.[1,3]

The internal oxidation mechanism has been applied to phenolics and has proved highly effective. Conventional wood-flour-filled phenolics have an arc resistance of 2 to 5 seconds, while conventional filled urea formaldehyde and melamine formaldehyde, which are normally felt to have good arc resistance, are in the range of 125 seconds.[14] With the internal oxidation mechanism the arc resistance of phenolics is increased from the previous 2 to 5 seconds to values of 180 to 190 seconds. Table III compares the ASTM arc resistance of conventional materials with phenolics using the internal oxidation mechanism.

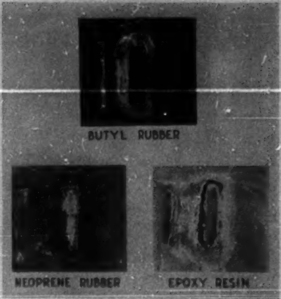


Fig. 10. Materials utilizing the internal oxidation mechanism after exposure to the dust-creepage test

Exploratory Work on Internal Ionization

When insulating materials are subjected to high-voltage stress, ionization will take place if internal voids are present and located in critical areas. Excessively severe internal ionization usually leads to rapid dielectric failure.[15] The internal oxidation mechanism, which has been shown to prevent surface carbonization, might also be expected to retard internal ionization and carbonization. Limited data on butyl rubber confirm the existence of this effect.

Transformers insulated with nontracking butyl are observed to have improved ionization characteristics after operating for a period of time at 300-per cent rated voltage. In addition, several nontracking butyl samples with molded-in electrodes retained over 50 per cent of their original

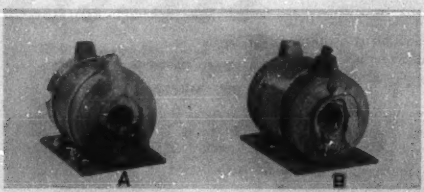


Fig. 11 (above). Conventional and nontracking epoxide-insulated 600-volt transformers after exposure to salt-fog test

A—Conventional epoxide, carbonized failure in less than 1 hour

B—Nontracking epoxide, no carbonization after 1,400 hours, failure by erosion

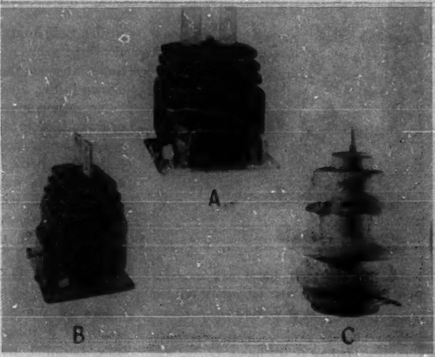


Fig. 12. Transformers and bushing after exposure to salt-fog test at 5 kv

A—Nontracking butyl transformer, no carbonization after 1,400 hours

B—Conventional butyl transformer, carbonized surface failure after 107 hours

C—Fifteen-kv porcelain bushing, cracked and chipped after 400 hours dielectric strength after internal failure on a 60-cycle high-potential test. This is apparently related to the absence of internal carbonization and to the formation of gas under pressure in the failure-path cavity.

An interesting experiment was conducted with a hypodermic needle inserted between the electrodes which confirmed the presence of gas under pressure after the initial breakdown. As long as the hypodermic needle was open to the atmosphere, the gas formed internally could not build up pressure, and the sample showed only the dielectric strength of air at atmospheric pressure. When the needle was closed to the atmosphere, the voltage could be raised slowly to over 50 per cent of the original breakdown value. By opening the hypodermic needle gas was released and the dielectric strength again dropped to the value for air. This cycle was repeated several times with the same results. A further study of this effect is progressing to obtain a better understanding of this phenomenon.

Summary

It has been discovered that the application of the internal oxidation mechanism to organic materials is highly effective in preventing carbonization under arcing conditions. Greatly improved resistance to surface-creepage damage has been observed in laboratory tests and under severe outdoor exposures. The formation of carbonized tracks is eliminated under these conditions. Standard ASTM arc-tracking tests indicate that phenolics using this mechanism can now be made superior to the best conventional arc-resistant molding compounds. The internal oxidation mechanism has also been observed to prevent the formation of conducting carbon following an internal dielectric failure.

Table III. ASTM 495-48T Arc-Resistance Test

Material	Average Time, Seconds
Phenolic	4
Urea	125
Melamine	125
Vulcanized fibre	125
Phenolic utilizing internal oxidation mechanism	185

References

1. Arc Resistance: (I) Tracking Processes in Thermosetting Insulating Materials, M. Olyphant, Jr. *ASTM Bulletin No. 181*, American Society for Testing Materials, Philadelphia, Pa., Apr. 1952, pp. 60–67.
2. Arc Resistance: (II) Effect of Testing Conditions on Tracking Properties of Thermosetting Insulating Materials, M. Olyphant, Jr. *ASTM Bulletin No. 185*, *Ibid.*, Oct. 1952, pp. 31–38.
3. Tracking in Solid Insulating Materials—A Critical Resume, V. E. Yarsley, W. J. Grant. *ERA Technical Report, Ref. L/T 177*, The British Electric and Allied Industry Research Association, London, England, 1948.
4. Tracking in Solid Insulating Materials—Consideration of Replies to Questionnaire, V. E. Yarsley. *ERA Technical Report, Ref. L/T 228*, *Ibid.*, 1950.
5. Insulation, H. Warren. *Journal*, Institution of Electrical Engineers, London, England, no. 528, 1940, pp. 588–96.
6. Vulcanized Fibre: New Look at an Old Plastic, G. Albert. *Materials and Methods*, New York, N. Y., Sept. 1954, pp. 110–14.
7. Application of Nontracking Butyl to High-Voltage Instrument Transformers, R. A. Pfuntner, R. S. Norman, B. W. Wilterdink. *AIEE Transactions*, vol. 74, pt. III, 1955, pp. 1173–76.
8. The Porcelain Insulator—Silent Partner on Power Transmission, H. A. Frey. *General Electric Review*, Schenectady, N. Y., Nov. 1954, pp. 52–56.
9. Insulator Surface Contamination, H. A. Frey. *AIEE Transactions*, vol. 67, pt. II, 1948, pp. 1420–25.
10. Flashovers of Suspension Insulators Due to Contamination, H. A. Adler, W. H. Wickham, M. S. Oldacre. *Ibid.*, pp. 1680–85.
11. Performance of Dust-Contaminated Insulators in Fog, B. Cozzens, T. M. Blakeslee. *Ibid.*, pp. 1686–92.
12. Arc-Resistance Test. *ASTM 495-48T*, American Society for Testing Materials, Philadelphia, Pa., 1948.
13. Letters to Insulation Forum, W. Vassar, C. Leff. *Insulation*, May 1955, p. 7.
14. Plastics Engineering Handbook, The Society of the Plastic Industry, Inc. Reinhold Publishing Corporation, New York, N. Y., 1954.
15. A New Dry-Type Insulation for Instrument Transformers, R. A. Pfuntner, R. E. Franck, F. R. D'Entremont. *AIEE Transactions*, vol. 69, pt. 1, 1950, pp. 453–58.

Transformer Reactance Calculations with Digital Computers

L. RABINS
ASSOCIATE MEMBER AIEE

SINCE one of the important features of a transformer is its series or leakage impedance, improvements in the calculation of this quantity are always searched for. In power transformers this impedance is almost all formed by its reactive component so that the accuracy of the impedance calculation hinges on the accuracy of the leakage reactance calculation. Modern methods of calculation using digital computers have been adopted so that very accurate reactance calculations can be made whenever they may be required in the design of a particular transformer. The use of a digital computer makes this calculation quick, easy, and painless, and permits a major improvement in this phase of transformer design.

In a cylindrical coil-core type of transformer, most of the difficulties in reactance calculation stem from axial ampere-turn discontinuities and unbalance between windings. The reactance of normal transformer designs in which both of these occur to a slight extent is accurately calculated by simple design equations. However, the design which may occur, for example, because of the request for extraordinary impedance relationships in a 3-winding transformer or as to parallel with a transformer having a peculiar impedance variation with the voltage tap position, leads to unusual axial ampere-turn distributions. This problem has been considered in the past,[1–5] but except for Roth's work[6] and unpublished work done within the General Electric Company by Liebinger, Poritsky, and Tino, there has been no consideration given the dual problem of axial ampere-turn distribution and the effect of curvature. It has become apparent that these two effects must be considered together and not separately, as in the past. Also in the past, investigators looked for simplification of rigorous equations. In the present work, the search has been for an accurate method adaptable to a digital-computer calculation.

Roth's analysis of this dual problem has followed in principle his double

---

Paper 56-59, recommended by the AIEE Computing Devices and Transformers Committees and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted September 29, 1955; made available for printing November 18, 1955.

L. Rabins is with the General Electric Company, Pittsfield, Mass.

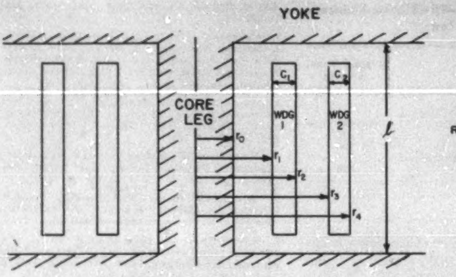


Fig. 1. Transformer with two cylindrical windings

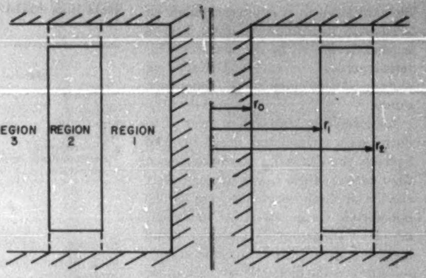


Fig. 2. Transformer with only one cylindrical winding shown

Fourier series method,[7] which he earlier applied to a reactance calculation in which curvature was neglected. With curvature considered, his analysis results in expressions difficult to evaluate not only because Bessel functions are involved but because the arguments of these functions are themselves functions of boundary dimensions and are found by solution of certain transcendental equations.

This paper is concerned with another and simpler form of solution to this problem, very similar in principle to Rogowski's[8] single Fourier series method, and classical methods of solution are employed.[9] The solution presented herein is in the form of a Fourier series the coefficients of which are Bessel and Struve functions, all of which are functions readily programmed for computation by an International Business Machines (IBM) card-programmed calculator.

The basic building block in the reactance calculation of any transformer is that between two windings. The reactance of any combination of windings can then be found by other methods, the most useful being that given by Garin and Paluev.[10] Only the basic 2-winding reactance calculation is being considered herein.

Outline of Problem and Method of Analysis

Fig. 1 shows a transformer with two cylindrical windings. The yokes above and below the coils are assumed of infinite extent and are completely circular so as to fully cover the windings. This permits the use of a Fourier series to describe the axial ampere-turn distribution of each winding. The general differential equations describing a magnetic field[11] are set up and the solution for the vector potential (which has only an angular component) is found. The magnetic field energy[11] can now be computed and the leakage inductance found.

Results and Calculations Procedure

The final equation used to compute reactance is

$$\%X=\frac{0.126f\text{ kva}}{e^2}\frac{(\Sigma NI)^2}{l}\times\left((Dr)+\frac{C_1^2-C_2^2}{12}\right)+\frac{0.063f\text{ kva }l}{e^2}\times\sum_{n=1}^{\infty}\left(\frac{I_{1n}(\text{pu})}{m^2C_1}\right)^2\left[\int_0^{x_2}xK_1dx\times\int_{x_1}^{x_2}xI_1dx+\left(\frac{I_0(x_0)}{K_0(x_0)}\int_{x_1}^{x_2}xK_1dx-\int_0^{x_1}xI_1dx\right)\int_{x_1}^{x_2}xK_1dx-\frac{\pi}{2}\int_{x_1}^{x_2}xL_1dx\right]+\left(\frac{I_{2n}(\text{pu})}{m^2C_2}\right)^2\times\left[\int_0^{x_4}xK_1dx\int_{x_3}^{x_4}xI_1dx+\left(\frac{I_0(x_0)}{K_0(x_0)}\int_{x_3}^{x_4}xK_1dx-\int_0^{x_4}xL_1dx\right)\times\int_{x_3}^{x_4}xK_1dx-\frac{\pi}{2}\int_{x_3}^{x_4}xL_1dx\right]+2\frac{I_{1n}(\text{pu})I_{2n}(\text{pu})}{m^4C_1C_2}\int_{x_3}^{x_4}xK_1dx\times\left(\frac{I_0(x_0)}{K_0(x_0)}\int_{x_1}^{x_2}xK_1dx+\int_{x_1}^{x_2}xI_1sx\right)\quad(1)$$

where

f = frequency, cycles per second
kva = kilovolt-ampere per core leg
e = volts per turn
l = axial length over which ampere-turn distribution is defined; will usually be either half or the whole distance between yokes
ΣNI = sum of per-unit (pu) ampere-turns in a coil stack
C_1, C_2 = radial builds of windings 1 and 2
Dr = effective area of leakage field for completely uniform flux field
$=[(r_2+r_1)/2](C_1/3)+[(r_3+r_2)/2]g+[(r_4+r_3)/2](C_2/3)$
g = radial gap between windings = r_3-r_2
n = index over which series is summed = 1, 2, 3, ...
$m=(n\pi)/l$
r_0, r_1, r_2, r_3, r_4 = radii defined by Fig. 1
$x=mr$
I_{1n}, I_{2n} = harmonics of the pu ampere-turn per inch distribution of windings 1 and 2
I_1, K_1, L_1 = modified Bessel and Struve functions of first order

The first term of the reactance equation results from the average ampere-turns in each winding, the average being taken over the length l. Flux lines corresponding to this reactance flow in a completely uniform vertical pattern over the entire length l and the reactance derivation is quite simple, appearing in several references,[12] so it will not be presented here. The remaining terms result from the nonuniform axial ampere-turn distribution. For most designs, symmetry exists about the axial centerline so that calculations may be made on a half-stack basis. If calculations are based on 1 pu ampere-turn in a full coil stack and the length l is half the distance between yokes, the same equation multiplied by 2 is used.

The harmonic coefficients of the ampere-turn distribution are given by (see Appendix I)

$$I_n=\frac{2}{n\pi}(\delta_{a-b}\sin ma+\delta_{b-c}\sin mb+\ldots\delta_{l-d}\sin md)\quad(2)$$

The calculation is performed on an IBM card-programmed calculator in a 10-digit floating decimal system. The calculation program is divided into two steps. First is the calculation of the harmonic coefficients. Here each winding is considered separately and broken into a minimum number of sections, each section having uniform ampere-turns pu axial height. The dimensions and total pu ampere-turns of each section are punched on IBM cards and are input data to the calculator. The program is arranged to handle any number of sections and, by omitting parts of the program deck, the length of the calculation is made proportional to the number of sections. After the first harmonic is evaluated, the integer n is advanced one, and an identical calculation is made, and this procedure is repeated until all necessary harmonics have been calculated. Harmonic coefficients are found for both windings in this fashion.

Following this, the actual reactance calculation is made. This requires additional input, frequency, kva per leg, volts per turn, and all radial dimensions beginning at the center line of the core and extending to the outer winding. The harmonic coefficients are also necessary input. These coefficients will have been punched on IBM cards during their calculation. Calculation then proceeds through the evaluation of the Bessel and Struve functions and the reactance, one harmonic at a time, each harmonic calculation procedure again being identical except for a change in harmonic order.

Derivation of Reactance Equation

The following is assumed in the derivation:

1. The yoke is infinitely large, completely covers top and bottom of the windings, and has infinite permeability.
2. Circular symmetry exists.
3. The effect of adjacent windings and core legs, tank walls, and current-carrying leads is negligible.
4. The field is unbounded radially.
5. The core leg is round and has infinite permeability.
6. Each winding is uniform radially.
7. The effect of normal ducts and insulation between coil sections is negligible.
8. The sum of the total ampere-turns in the two windings is zero.

The field equations appropriate to this problem of reactance calculation are

$$\nabla\cdot\mathbf{B}=0$$

$$\nabla\times\mathbf{H}=\mathbf{i}$$

$$\mathbf{B}=\mu\mathbf{H} \tag{3}$$

where **B**, **H**, and **i** are flux density, magnetic field strength, and current density respectively. Meter-kilogram-second units are used. Define a vector potential by

$$\nabla\times\mathbf{A}=\mathbf{B}$$

$$\nabla\cdot\mathbf{A}=0 \tag{4}$$

Substituting the foregoing results in[11]

$$\nabla^2\mathbf{A}=-\mu\mathbf{i} \tag{5}$$

The vector **i** defines the current density in the winding and since it has only an angular component, this will also be the only component of the vector potential. Denoting this component by the subscript

$$\nabla^2 A_\theta=-\mu i_\theta \tag{6}$$

In circular co-ordinates this becomes

$$\frac{\partial^2 A_\theta}{\partial r^2}+\frac{1}{r}\frac{\partial A_\theta}{\partial r}-\frac{A_\theta}{r^2}+\frac{\partial^2 A}{\partial z^2}=-\mu i_\theta \tag{7}$$

Describe the current density by a Fourier series (see Appendix I)

$$i_\theta=i_0+\sum_{n=1}^{\infty} i_n\cos mZ \tag{8}$$

where $n=1, 2, 3, \ldots$. Assume that the vector potential can be similarly developed

$$A_\theta=A_{0(r)}+\sum_{n=1}^{\infty} A_{n(r)}\cos mZ \tag{9}$$

Substituting the foregoing, a set of equations is obtained

$$\frac{d^2A_0}{dr^2}+\frac{1}{r}\frac{dA_0}{dr}-\frac{A_0}{r^2}=-\mu i_0 \tag{10}$$

$$\frac{d^2A_n}{dr^2}+\frac{1}{r}\frac{dA_n}{dr}-\frac{A_n}{r^2}-m^2A_n=-\mu i_n \tag{11}$$

for each n. This set of equations is obtained because of the orthogonality of $\cos mZ$. The first of these equations is easy to solve and corresponds to flux flowing in only the Z-direction. The reactance resulting from this component of the vector potential is the first term in the reactance equation previously given in the section entitled "Results and Calculation Procedure." It is the set of equations for general n which is mostly of interest. If the right-hand side of the foregoing equation were zero, the equation would be recognized as Bessel's and the solution would be[13]

$$A_n=CI_1(mr)+DK_1(mr) \tag{12}$$

where I_1 and K_1 are modified Bessel functions of the first and second kind, first order. A particular solution to the complete equation may be found by the method of quadratures[16] or by recognition that it is satisfied by a modified Struve function $L_1(mr)$.[13–15] The complete solution for general n is then

$$A_n=CI_1(mr)+DK_1(mr)-\frac{\pi\mu i_n}{2m^2}L_1(mr) \tag{13}$$

C and D are constants determined by boundary conditions. Consider one winding at a time and the three regions defined by Fig. 2. The solutions to the differential equations in the three regions are given respectively as

$$A_{1n}=C_1I_1(mr)+D_1K_1(mr)$$

$$A_{2n}=C_2I_1(mr)+D_2K_1(mr)-\frac{\pi\mu i_n}{2m^2}L_1(mr)$$

$$A_{3n}=C_3I_1(mr)+D_3K_1(mr) \tag{14}$$

The boundary conditions reveal information which permit evaluation of the constants C and D. These stem from continuity of the flux densities. From the definition of the vector potential

$$B_r=-\frac{\partial A_\theta}{\partial Z}$$

$$B_Z=\frac{1}{r}\frac{\partial rA}{\partial r} \tag{15}$$

At the iron boundaries, the tangential flux is zero since infinite permeability iron was assumed. Thus

$$B_r=0 \text{ at } Z=\pm l \tag{16}$$

This yields the equation

$$\sum_{n=1}^{\infty}-mA_n\sin ml=0$$

or

$$m=\frac{n\pi}{l}, n=1, 2, 3, \ldots \tag{17}$$

The other boundary conditions are

$$B_Z=0 \text{ at } r=r_0$$

$$B_{1Z}=B_{2Z};\ B_{1r}=B_{2r} \text{ or } A_{1n}=A_{2n} \text{ at } r=r_1$$

$$B_{2Z}=B_{3Z};\ B_{2r}=B_{3r} \text{ or } A_{2n}=A_{3n} \text{ at } r=r_2$$

$$B_Z \text{ or } A_Z\rightarrow 0 \text{ as } r\rightarrow\infty \tag{18}$$

The last condition requires that

$$C_3=0 \tag{19}$$

since $I_1(mr)$ is an exponentially increasing function.

For any value n, the general value of B_{Zn} is

$$B_{Zn}=m\left(CI_0(mr)-DK_0(mr)-\frac{\pi\mu i_n}{2m^2}L_0(mr)\right) \tag{20}$$

Let $mr=x$ and

$$-\frac{\pi\mu i_n}{2m^2}=i_n' \tag{21}$$

Substitute the remaining boundary conditions in the equations 14 and 20 for A_n and B_{Zn} to obtain

$$C_1I_0(x_0)-D_1K_0(x_0)=0$$

$$C_1I_1(x_1)+D_1K_1(x_1)=C_2I_1(x_1)+D_2K_1(x_1)+i_n'L_1(x_1)$$

$$C_1I_0(x_1)-D_1K_0(x_1)=C_2I_0(x_1)-D_2K_0(x_1)+i_n'L_0(x_1) \quad (22)$$

$$D_3K_1(x_2)=C_2I_1(x_2)+D_2K_1(x_2)+i_n'L_1(x_2)$$

$$-D_3K_0(x_2)=C_2I_0(x_2)-D_2K_0(x_1)+i_n'L_0(x_2)$$

Utilize the equations

$$I_0(x)K_1(x)+I_1(x)K_0(x)=\frac{1}{x} \text{ (see reference 13)}$$

$$x[L_1(x)K_0(x)+L_0(x)K_1(x)]=\frac{2}{\pi}\int_0^x xK_1(x)dx$$

$$x[L_1(x)I_0(x)-L_0(x)I_1(x)]=\frac{-2}{\pi}\int_0^x xI_1(x)dx \quad (23)$$

The origin of the latter equations is covered in Appendix II. Solving for C and D

$$C_1=\frac{\mu i_n}{m^2}\int_{x_1}^{x_2} xK_1dx$$

$$C_2=\frac{\mu i_n}{m^2}\int_0^{x_2} xK_1dx$$

$$D_1=\frac{I_0(x_0)}{K_0(x_0)}\,\frac{\mu i_n}{m^2}\int_{x_1}^{x_2} xK_1dx \quad (24)$$

$$D_2=\frac{I_0(x_0)}{K_0(x_0)}\,\frac{\mu i_n}{m^2}\int_{x_1}^{x_2} xK_1dx-\frac{\mu i_n}{m^2}\int_0^{x_1} xI_1dx$$

$$D_3=\frac{I_0(x_0)}{K_0(x_0)}\,\frac{\mu i_n}{m^2}\int_{x_1}^{x_2} xK_1dx+\frac{\mu i_n}{m^2}\int_{x_1}^{x_2} xI_1dx$$

The solution for the vector potential is now complete.

Consider now that two windings are present, as shown in Fig. 1. Two solutions of the same form as that just found may be superimposed by properly identifying regions 1, 2, and 3. The space harmonics of the ampere-turn distribution of windings 1 and 2 will be represented by i_{1n} and i_{2n} respectively. In the region occupied by winding 1

$$A_{1n}=\frac{\mu i_{1n}}{m^2}\Bigg[I_1(x)\int_0^{x_2} xK_1dx+K_1(x)\times\left(\frac{I_0(x_0)}{K_0(x_0)}\int_{x_1}^{x_2} xK_1dx-\int_0^{x_1} xI_1dx\right)-\frac{\pi}{2}L_1(x)\Bigg]+\frac{\mu i_{2n}}{m^2}\left(I_1(x)\int_{x_3}^{x_4} xK_1dx+K_1(x)\frac{I_0(x_0)}{K_0(x_0)}\int_{x_3}^{x_4} xK_1dx\right) \quad (25)$$

In the region occupied by winding 2

$$A_{2n}=\frac{\mu i_{2n}}{m^2}\Bigg[I_1(x)\int_0^{x_4} xK_1dx+K_1(x)\times\left(\frac{I_0(x_0)}{K_0(x_0)}\int_{x_3}^{x_4} xK_1dx-\int_0^{x_3} xI_1dx\right)-\frac{\pi}{2}L_1(x)\Bigg]+\frac{\mu i_{1n}}{m^2}K_1(x)\left(\frac{I_0(x_0)}{K_0(x_0)}\times\int_{x_1}^{x_2} xK_1dx+\int_{x_1}^{x_2} xI_1dx\right) \quad (26)$$

REACTANCE DERIVED FROM VECTOR POTENTIAL

With the vector potential known, the leakage inductance may be found by evaluating the energy stored in the magnetic field. If W is the magnetic field energy, and L the inductance[11]

$$W=\frac{1}{2}LI^2=\frac{1}{2}\iiint \bar{\mathrm{i}}\cdot\bar{\mathrm{A}}dv \quad (27)$$

Both $\bar{\mathrm{i}}$ and $\bar{\mathrm{A}}$ have only angular components each given by a Fourier series. dv is a differential volume $=rd\theta dZdr$.

Thus

$$\iiint \mathrm{i}\cdot\bar{\mathrm{A}}dv=\iiint\left(i_0+\sum_{n=1}^{\infty}i_n\cos mZ\right)\times\left(A_0+\sum_{n=1}^{\infty}A_n\cos mZ\right)dv \quad (28)$$

In the integration all terms not of the same order go to zero since

$$\int_0^l \cos m_1Z\cos m_2Z\,dZ=l/2 \quad (29)$$

if $m_1=m_2$, or equals zero if $m_1\neq m_2$.

This leaves

$$\iiint \mathrm{i}\cdot\mathbf{A}dv=l\iint i_0A_0rd\theta dr+\sum_{n=1}^{\infty}l/2\iint i_nA_nrd\theta dr \quad (30)$$

Carrying the integration over only the windings since it is only here that $i\neq0$ and omitting the first term in the foregoing

$$L=\frac{\iiint \mathrm{i}\cdot\mathbf{A}dv}{I^2}=\sum_{n=1}^{\infty}\frac{l}{2I^2}\int_0^{2\pi}\int_{r_1}^{r_2} i_{1n}A_{1n}rd\theta dr+\int_0^{2\pi}\int_{r_3}^{r_4} i_{2n}A_{2n}rd\theta dr \quad (31)$$

Except for the integral of the Bessel functions, a typical term of this expression has the coefficient

$$\mathrm{coef}_L=\frac{\pi l\mu i_{1n}i_{2n}}{m^4I^2} \quad (32)$$

Refer both windings to ones with an equal number of turns N. Then, with the current density being uniform over the radial build, it may be written as

$$i_{1n}=\frac{I_{1n}}{C_1} \quad (33)$$

where I_{1n} is the ampere-turns pu axial length. Then

$$\mathrm{coef}_L=\frac{\pi l\mu I_{1n}I_{2n}}{C_1C_2m^2I^2} \quad (34)$$

Multiplying numerator and denominator by N^2 and recognizing

$$I_{1n}/(NI)=I_{1n}(\mathrm{pu})=\text{pu ampere-turns pu axial length} \quad (35)$$

$$\mathrm{coef}_L=\frac{\pi l\mu N^2I_{1n}(\mathrm{pu})I_{2n}(\mathrm{pu})}{C_1C_2m^4} \quad (36)$$

This is the coefficient of an expression for the leakage inductance; to change to per-cent reactance multiply by

$$2\pi f\,10^5\frac{\mathrm{kva}}{V^2} \quad (37)$$

where

V = the rated voltage of a winding
kva = kva rating of a winding
f = frequency

The foregoing coefficient has a length dimension to the first power; changing from meters to inches and using the permeability of free space of $4\pi\times10^{-7}$

$$\mathrm{coef}_{\%X}=0.063\frac{f\,\mathrm{kva}\,l}{(V/N)^2}I_{1n}(\mathrm{pu})I_{2n}(\mathrm{pu}) \quad (38)$$

The final expression is now found by carrying out the integration in equation 31 for reactance and is given as equation 1.

Bessel Function Computation

The general process by which the Bessel and Struve functions were evaluated on the card-programmed calculator consisted of a computation using either a power or asymptotic series.

The card program is set up so that for low arguments a power series is used and for high arguments an asymptotic series. The changeover point is 6 for I_0, $\int xI_1dx$, and $\int xL_1dx$; and 3.5 for K_0 and $\int xK_1dx$. In general, 12 terms were used in the power series expansion, 7 or less in the asymptotic. To obtain a smooth transition of the integrals at the changeover point, it was necessary to adjust the last coefficient of the asymptotic series.

The general principle involved in the computation of the series was built on adding terms of the form

$$xC+D$$

each such operation being performed in one card cycle. A power series with seven terms

$$P=A_0+A_1x+A_2x^2+A_3x^3+A_4x^4+A_5x^5+A_6x^6$$

may be factored starting at the end to give

$$P=A+x[A_1+x(A_2+x\{A_3+x[A_4+x(A_5+xA_6)]\})]$$

so that it is seen that one card operation advances one term in the evaluation of the series. Based on a speed of 150

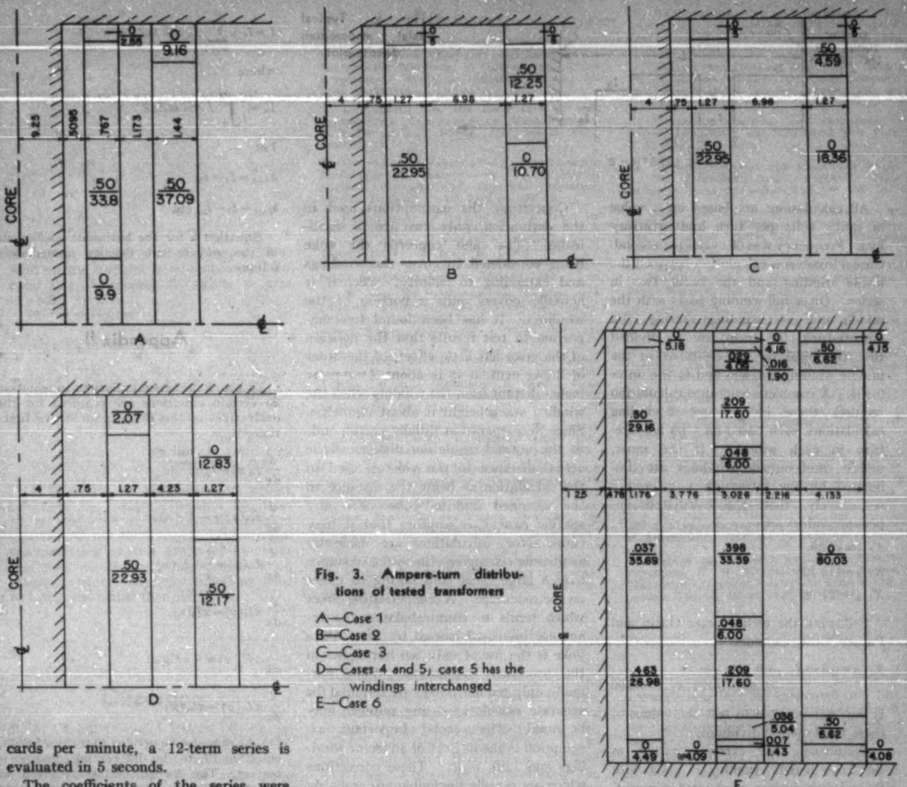

CORE
A
9.25
.5095
.767
1.173
1.44
0 / 2.55
0 / 9.16
.50 / 33.8
.50 / 37.09
0 / 9.9
B
4
.75
1.27
5.98
1.27
0 / 5
.50 / 12.25
.50 / 22.95
0 / 10.70
C
.50 / 4.59
.50 / 22.95
0 / 18.36
D
4.23
0 / 2.07
0 / 12.83
.50 / 22.93
.50 / 12.17
E
1.25
.478
1.176
3.776
3.026
2.216
4.133
0 / 5.16
.50 / 29.16
.037 / 35.69
.463 / 26.98
0 / 4.49
0 / 4.09
0 / 4.16
.029 / 4.09
.016 / 1.90
.209 / 17.60
.048 / 6.00
.398 / 33.59
.036 / 5.04
.007 / 1.43
.50 / 6.62
0 / 4.15
0 / 80.03
0 / 4.08

Fig. 3. Ampere-turn distributions of tested transformers

A—Case 1
B—Case 2
C—Case 3
D—Cases 4 and 5; case 5 has the windings interchanged
E—Case 6

cards per minute, a 12-term series is evaluated in 5 seconds.

The coefficients of the series were evaluated by a desk calculator using general expressions.[13,14] The series for the integrals were found by direct integration of the power series and by repeated integration by parts of the asymptotic series.

The functions were calculated over the range 0 to 50 and compared to existing tables,[17–20] or to combinations of existing tables when direct comparison could not be done. Better than 5-place accuracy was obtained throughout.

It should be noted that considerable simplicity results in the determination of the series of the functions to be evaluated from the use of Struve functions and the recognition of the integrals which are equal to certain groups of Bessel and Struve functions. If the Struve functions were not used as a solution of the differential equation, but a solution built up by superposition and summation of the effect of current sheets, or if the solution had been left in the integral form given by the method of quadratures, the result for the reactance would have contained various combinations of terms such as

$$\int_{x_1}^{x_2} xK_1dx\int_{x_1}^{x} xI_1dx$$

which is not in a form in which the coefficients of the power or asymptotic series are readily found.

Test Results

Several comparisons between tests and calculated reactances on production transformers and models which were built to get an exaggerated ampere-turn distribution are listed in the following. These examples are all normally difficult to calculate. Fig. 3 shows the general winding configurations. The data within each block on the figures contain the pu ampere-turns (the top number) and the axial height of the block (the bottom number). Distances to the yoke and holes in a winding are shown as blocks with zero ampere-turns. All values of pu ampere-turns are based on a total of 1 pu ampere-turn in a stack. Table I summarizes the test results. In all cases the calculation was made through the tenth harmonic.

Table I. Test Results

Case	Kva	Per-Cent X	
		Test	Calculated
1	1.63	14.5	13.75
2	3.60	58.7	58.29
3	3.60	97.6	95.53
4	3.60	49.5	42.10
5	3.60	43.5	43.66
6	2.53	34.3	35.99

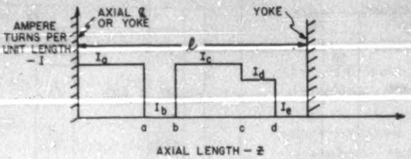


Fig. 4. Typical axial ampere-turn density plot

All calculations are based on a value of unity volts per turn and arbitrary kva. Frequency was 60 cycles per second. Case 6 involves a calculation between the inside winding and the other two in series. On a full-winding basis with the middle and outer windings sharing 1 pu ampere-turn, the turns are so divided that the total pu ampere-turns in the middle winding is 0.896 and in the outer 0.104. A complete reactance calculation requires three independent 2-winding calculations, each based on 1 pu ampere-turn in each winding. If the inner, middle, and outside windings are designated by the subscripts 1, 2, and 3 respectively, then the calculated impedance values are

$X_{1-2}=40.39$

$X_{1-3}=112.42$

$X_{2-3}=127.76$

Following the technique of Garin and Paluev,[10] the reactance

$$X_{1-23}=0.896X_{1-2}+0.104_{1-3}-(0.896)(0.104)X_{2-3}=35.99$$

It was not possible to test the values of these reactances individually.

Considering the extreme nature of the cases listed, accuracy is good with the maximum error of about 5 per cent. Some of this error may be due to manufacturing tolerances or slight errors in test.

Discussion and Conclusions

Equations and computational techniques for the calculation of the leakage reactance of core-type cylindrical transformer windings have been presented. While emphasis has been placed on the use in transformer design, the results are applicable to air-core reactors also.

The interpretation of the reactance equation is straightforward. Each harmonic contribution is seen to consist of two self-reactances terms and a mutual reactance term. The reactances are only that in a leakage field sense, i.e., they do not include the contribution of the core. This avoids the difficulty of finding a small leakage reactance by taking differences of very large reactances.

Concerning the assumptions used in the derivation, only two are of significance. The first concerns the yoke being considered to cover the windings and extending to infinity; whereas it actually covers only a portion of the windings. It has been found by comparison to test results that the position of the yoke has little effect, of the order of 3 per cent, if it is about 4 or more inches distant from the winding when the winding stack height is about 90 inches. Since this separation usually corresponds to the normal insulation distances, the actual distance to the yoke is used in the calculation as being the distance to the assumed infinite yoke. For any specific case, if it appears that it may cause error, calculations are made by arbitrarily changing the yoke distance over a large range and noting its effect on the reactance. A compensating effect which tends to counterbalance the reactance increased caused by the infinite yoke is the use of only ten harmonics in the reactance calculation. While normally only ten harmonics are required for accurate calculations, any number may be used. The second important assumption is the neglect of adjacent windings and tank walls. These corrections which are usually negligible are probably best handled empherically, though some qualitative effect may be found by locating a highly permeable or highly conducting cylinder outside the windings and by determining the reactances for these cases. The reactance equations for these conditions may be found by the same methods as described herein.

It should be emphasized that the one major item which makes the use of these accurate methods of reactance equations practical for regular design application is the modern high-speed digital computer.

Appendix I

A plot of the ampere-turns per unit axial height versus height is a series of straight-line sections, as shown in Fig. 4. Over the length l the ampere-turn density may be defined by a Fourier series

$$I=I_0+\sum_{n=1}^{\infty} I_n \cos mZ;\; m=\frac{n\pi}{l}$$

where

$$I_n=\frac{2}{l}\int_0^l I \cos mZdZ$$

Let

$\delta_{a-b}=I_a-I_b$

$\delta_{b-c}=I_b-I_c$, etc.

Equation 2 for the harmonic coefficients of the ampere-turn density immediately follows.

Appendix II

To prove the several relations in equation 23 certain equations are required for the derivatives of the Bessel and Struve functions[14]

$$\frac{d}{dx}I_0(x)=I_1(x)$$

$$\frac{d}{dx}K_0(x)=-K_1(x)$$

$$\frac{d}{dx}L_0(x)=\frac{2}{\pi}+L_1(x)$$

$$\frac{d}{dx}xI_1(x)=xI_0(x)$$

$$\frac{d}{dx}xK_1(x)=-xK_0(x)$$

$$\frac{d}{dx}xL_1(x)=xL_0(x)$$

It may now also be verified that the solution, equation 13, to the differential equation is correct. The integral relations given in equation 23 were first obtained by noting a general equation given by McLachlan and Meyers[21] for

$\int x^n I_n(x)dx$ and $\int x^n K_n(x)dx$

and by comparing the equation thereby obtained for

$\int I_0(x)dx$ and $\int K_0(x)dx$

to an integration by parts of

$\int xI_1(x)nx$ and $\int xK_1dx$

This procedure does not have to be followed here since the integrals may be verified by using the derivative equations in the foregoing and noting that

$$\frac{d}{dx}(xL_1(x)K_0(x)+xL_0(x)K_1(x))=\frac{2}{\pi}xK_1(x)$$

$$\frac{d}{dx}(xL_1(x)I_1(x)-xL_0(x)I_1(x))=-\frac{2}{\pi}xI_1(x)$$

By making use of the facts that for small values of x

$I_0(x)=1$

Vol.
Pt. 1
1956

$$K_0(x) = -ln\frac{x}{2}$$

$$L_0(x) = \frac{2x}{\pi}$$

$$I_1(x) = \frac{x}{2}$$

$$K_1(x) = \frac{x}{2}\, ln\frac{x}{2} + \frac{1}{x}$$

$$L_1(x) = \frac{2x}{3\pi}$$

it may be verified that the proper lower limit on the integral is zero since the left-hand side of equation 23 reduce to zero at $x=0$.

This is shown as

$$\{x[L_1(x)K_0(x) + L_0(x)K_1(x)]\}_{x=0}$$

$$= \lim_{x\to 0} x\left[\frac{-2x}{3\pi}\, ln\frac{x}{2} + \frac{2x}{\pi}\left(\frac{x}{2}\, ln\frac{x}{2} + \frac{1}{x}\right)\right] = 0$$

$$\{x[L_1(x)I_0(x) - L_0(x)I_1(x)]\}_{x=0}$$

$$= \lim_{x\to 0} x\left(\frac{2x}{3\pi} - \frac{x^2}{\pi}\right) = 0$$

It is necessary to determine proper limits on the integrals because the term to be evaluated is the left-hand side of equation 23, not the integral. The integral representation of this group of functions was used to obtain a faster calculation procedure since only two functions need be calculated rather than six.

References

1. TRANSFORMER REACTANCE AND LOSSES WITH NONUNIFORM WINDINGS, H. O. **Stephens.** *AIEE Transactions (Electrical Engineering)*, vol. 53, 1934, pp. 346–49.

2. A FORMULA FOR THE REACTANCE OF THE INTERLEAVED COMPONENT OF TRANSFORMERS, **H. B. Dwight, L. S. Dzung.** *Ibid.*, vol. 56, 1937, pp. 1368–71.

3. INFLUENCE OF VARIOUS FACTORS ON LEAKAGE REACTANCE OF TRANSFORMERS, **A. L. Morris.** *Journal*, Institution of Electrical Engineers, London, England, vol. 86, 1940, pp. 485–95.

4. CALCULATION OF THE MAGNETIC FIELD OF RECTANGULAR CONDUCTORS IN A CLOSED SLOT AND ITS APPLICATION TO THE REACTANCE OF TRANSFORMER WINDINGS, **E. Billig.** *Report Q/T106*, British Electrical and Allied Industries Research Association, London, England, 1945.

5. ON THE NEW METHOD OF TRANSFORMER LEAKAGE REACTANCE CALCULATION, **K. Ozawa.** *Japan Science Review*, Tokyo Institute of Technology, Tokyo, Japan, vol. 1, Mar., 1949, pp. 69–75.

6. MAGNETIC LEAKAGE INDUCTANCE IN TRANSFORMERS WITH CYLINDRICAL WINDINGS AND FORCES ACTING ON THE WINDINGS, **E. Roth.** *Revue Générale de L'Électricité*, Paris, France, vol. 40, 1936, pp. 291–303, 323–36.

7. ELECTROMAGNETIC PROBLEMS IN ELECTRICAL ENGINEERING (book), **B. Hague.** Oxford University Press, Oxford, England, 1929, pp. 298–325.

8. ON THE LEAKAGE FIELD AND THE LEAKAGE INDUCTION COEFFICIENT OF A TRANSFORMER WITH SANDWICH WINDINGS AND DIVIDED END COILS, **W. Rogowski.** *Mitteilungen über Forschungsarbeiten*, Zeitschrift des Vereins Deutscher Ingenieure, Düsseldorf, Germany, vol. 71, 1909, pp. 1–36.

9. STATIC AND DYNAMIC ELECTRICITY (book), **W. R. Smythe.** McGraw-Hill Book Company, Inc., New York, N. Y., first edition, 1939, pp. 290–93.

10. TRANSFORMER CIRCUIT IMPEDANCE CALCULATIONS, **A. N. Garin, K. K. Paluev.** *AIEE Transactions (Electrical Engineering)*, vol. 55, June 1936, pp. 717–20.

11. STATIC AND DYNAMIC ELECTRICITY (book), **W. R. Smythe.** McGraw-Hill Book Company, Inc., New York, N. Y., first edition, 1939, pp. 256–322.

12. MAGNETIC CIRCUITS AND TRANSFORMERS (book), John Wiley and Sons, Inc., New York, N. Y., 1943, pp. 358–62.

13. BESSEL FUNCTIONS FOR ENGINEERS (book), **N. W. McLachlan.** Oxford University Press, Oxford, England, first edition, 1934.

14. THE STER AND STEI FUNCTIONS, **N. W. McLachlan, A. L. Meyers.** *Philsophical Magasine*, Taylor and Francis, Ltd., London, England, vol. 21, 1936, pp. 425–36.

15. THEORY OF BESSEL FUNCTIONS (book), **G. N. Watson.** Cambridge University Press, Cambridge, England, 1922, pp. 328–38.

16. DIFFERENTIAL EQUATIONS (book), **L. R. Ford.** McGraw-Hill Book Company, Inc., New York, N. Y., p. 76.

17. **BRITISH ASSOCIATION MATHEMATICAL TABLES.** Cambridge University Press, Cambridge, England, vol. 6, BESSEL FUNCTIONS.

18. TABLE OF THE FUNCTIONS, $\int K_0 dx \int I_0 dx$, **V. R. Bursian, V. Fock.** *Travaux*, Institute of Physics and Mathematics, Leningrad, USSR, vol. 2, 1931, pp. 1–10.

19. TABLES OF THE STRUVE FUNCTIONS, MATHEMATICAL TABLES PROJECT, **National Bureau of Standards.** *Journal of Mathematics and Physics*, Massachusetts Institute of Technology, Cambridge, Mass., vol. 25, 1946, pp. 252–59.

20. TABLES OF INTEGRALS OF STRUVE FUNCTIONS, **M. Abramowitz.** *Ibid.*, vol. 29, 1950, pp. 49–51.

21. INTEGRALS INVOLVING BESSEL AND STRUVE FUNCTIONS, **N. W. McLachlan, A. L. Meyers.** *Philosophical Magasine*, Taylor and Francis, Ltd., London, England, vol. 21, 1936, pp. 437–48.

Discussion

B. A. Cogbill (General Electric Company, Pittsfield, Mass.): The title of Mr. Rabins' paper might imply that he has done a more or less routine job of adapting a method of computing transformer reactance to digital computer calculations. However, it is obvious that this is a minor part of his accomplishment; more important is his mathematical analysis of the field problem.

It may be correctly argued that the field problem has been analzyed many times in the past by very capable investigators. But a thorough study of published material reveals that, in practically all cases, the earlier investigators were looking for simplification. Digital computers were unknown. So earlier efforts were directed toward obtaining expressions which were practical for use in calculating transformer reactance by methods available at the time. Hence, a number of simplifying assumptions were usually injected at an early stage of the analysis; the almost universal one being that curvature would be ignored. As a result, the use of the expressions for reasonably accurate calculations is quite restricted; and, even in the simpler cases, results are somewhat less accurate than those which could be obtained from use of more rigorous expression.

No doubt Mr. Rabins considered the existing expressions and their derivations. But knowing the capabilities of digital computers, he was confident that they could handle with comparative ease the more complicated calculations involved in the use of more rigorous methods. So, he chose the more difficult, but more rewarding, course of analyzing the field problem to obtain a solution which met his needs. This was accomplished in an admirable manner. The mathematical analysis of the field problem alone is an outstanding contribution and should be of interest to many engineers and practical mathematicians although they might not have any particular interest in digital computers.

The paper contains a number of comparisons between test values and values calculated on the digital computer using the author's method. As stated, these cases are of extreme nature. The author cited them to demonstrate the accuracy of the method under conditions for which other methods are known to be quite unreliable. But he had no intention of conveying the impression that the method is not practical for use in the more usual cases. Actually, a large number of reactances for power transformer designs have been calculated by the method; and it has proved to be a practical design tool.

The amount of data available for making a comparison between calculated and test values is much larger now than it was at the time the paper was written. Therefore a more realistic appraisal can be made by using only reactances of transformers which have been built and are in service. Based on nine reactances which have been calculated by the method under consideration and tested after final assembly of the transformers, the average of the ratio of calculated value to test value is 0.978; and one standard deviation from this average is only 0.0101. This is quite impressive, especially when it is considered that some of the error may be due to manufacturing tolerances and slight inaccuracies in test. Moreover, none of the transformer designs on which the method has been used is of the simplest type where standard equations provide us with sufficiently accurate results.

L. Rabins: I wish to thank Mr. Cogbill for his discussion and for the presentation of the data on production transformers. As he has stated, existing rigorous methods for reactance calculation were investigated. Calculations on the card-programmed calculator showed these to yield insufficiently accurate results so that the need for a less restrictive approach became apparent.

Flux Reversal in Magnetic Amplifier Cores

F. J. FRIEDLAENDER
ASSOCIATE MEMBER AIEE

IN RECENT years theun derstanding magnetic amplifier circuits has advanced rapidly, and great strides have been made in the art. However, it is recognized that several aspects of magnetic amplifier operation still need clarification. In most instances, the lack of a satisfactory explanation for certain circuit characteristics can be traced to an incomplete understanding of the behavior of the magnetic cores used. Therefore, a definite need exists for a study of the process of flux reversal in these magnetic amplifier components. The grain-oriented 50-per-cent nickel-iron alloys (such as Orthonol) are of particular interest and will be the material considered here. However, most of the results can probably be extended directly to other highly oriented "square hysteresis loop" materials.

A convenient starting point is found in an area in which considerable work has already been done, both experimentally and theoretically; namely, flux reversal under the influence of a step function of the magnetomotive force (mmf). This situation arises practically in many magnetic amplifier circuits and is closely approximated in magnetic amplifier applications with high control-circuit impedance. In the computer field similar situations are also encountered, though a less detailed knowledge of the process is usually required there.

The application of a step of mmf to a magnetic core, in addition to its practical importance, provides a forcing function comparatively easy to handle, both from a mathematical and an experimental point of view. As a basic procedure, the voltage induced in a separate pickup coil on the core has been observed upon application of a step of mmf, as done by previous experimenters.[1,2,3] Experimental evidence has been compared at all steps with theoretical considerations developed and modified as the work progressed. In the course of the study, conceptions expressed by physicists on the nature of ferromagnetism have been considered carefully; on the whole, however, the effort has been directed towards the development of an engineering model by engineering methods to meet immediate needs.

The model thus developed is utilized further as a basis for the study of other types of flux reversal, as occurring upon the application of other forcing functions, i.e., sinusoidal voltage. The ability of the model to account for these processes serves as a check on its usefulness.

Past Concepts of Flux Reversal in Tape Cores

The cores employed in magnetic amplifiers usually consist of a spiral of thin nickel-iron tape. It is convenient to begin an analysis of the magnetization processes in such cores by considering a single layer of the tape formed into a closed, circular cylinder with a radius r which is very large compared with the thickness d of the tape. A magnetic field is provided by a current flowing in an exciting winding placed toroidally on the tape. Then for a current of i amperes flowing in a winding of N turns, wound on a tape of length l centimeters where $l = 2\pi r$, considered constant over the cross section of the tape since $d << r$, the magnetic intensity at the tape is

$$\frac{Ni}{l}\ \frac{\text{ampere-turns}}{\text{centimeters}} = 0.4\pi \frac{Ni}{l}\ \text{oersted}$$

The voltage induced in a separate pickup coil, placed on the same tape core, has to be found as a function of time when a constant current I is suddenly applied to the exciting winding.

The original approach to the problem was to solve Maxwell's equations for the given geometry, with suitable assumptions about the B-versus-H relationship (i.e., the value of μ). If a linear B-versus-H relationship is assumed, an easily solved boundary value problem results.[4] However, it cannot be expected that such a treatment will give reasonably accurate predictions of the induced voltage (i.e., the rate of change of flux) during flux reversal for materials having very nonlinear B-versus-H relationships, such as a rectangular hysteresis loop. This approach may still be useful in predicting induced voltages occurring above the knee of a rectangular hysteresis loop in a saturated core.

It was recognized early that a realistic analysis would require careful consideration of the nonlinearities involved. Many approximate mathematical treatments were attempted. Among them are those considering a sectionally linear hysteresis loop. As a limiting case of such a loop, a rectangular B-versus-H relationship was assumed, ($\mu = \infty$ or $\mu = 0$) leading to a simpler mathematical formulation.

Wolman and Kaden[4] obtained the solution to this case, modified by the addition of a hysteresis term, with the use of the hysteresis loop of Fig. 1 under the following assumptions:

1. The entire tape is initially saturated.
2. Flux reversal takes place along a boundary of zero thickness.
3. Flux reversal starts in the two planes, represented by the large surfaces of the tape, and propagates into the material; see Fig. 2.
4. The value of H at the plane of flux reversal is equal to H_c, as required by the assumed hysteresis loop.

In this analysis, with a step of current I applied at $t=0$ resulting in a step of mmf denoted by F_k, the induced voltage per turn is found to be

$$\frac{e}{N} = -2w\sqrt{\frac{\rho H_s B_s}{0.4\pi t}}$$

where

$$H_s = 0.4\pi \frac{NI}{l} - H_c \qquad (1)$$

The time T_s required for flux reversal is also of interest. It is known as the switching time in the computer art

$$T_s = \frac{0.4\pi d^2}{4\rho H_s} \qquad (2)$$

More recently Papoulis[5] obtained essentially identical results to those of Wolman and Kaden[4] by more involved techniques for the solution of Maxwell's equations under the foregoing conditions.

Since the postulation of the modern theory of ferromagnetism by Pierre Weiss in 1907, it has been recognized that a ferromagnetic material consists of regions, known as domains, in which all the magnetic dipoles are aligned even without application of an external field.[6,7,8] In

---

Paper 56-219, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing December 27, 1955.

F. J. FRIEDLAENDER is with Purdue University, Lafayette, Ind.

This work was performed in part under sponsorship of the United States Office of Naval Research. Portions of the paper were abstracted from a dissertation by F. J. Friedlaender, submitted to Carnegie Institute of Technology in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author gratefully acknowledges the advice and comments of Professors Leo A. Finzi and Jacob E. Goldman with the Ford Motor Company, and Howard M. McConnell, now with Jack and Heintz. Dr. L. J. Dijkstra, Westinghouse Research Laboratories, and Harold Lord, General Electric Research Laboratories, are thanked for their stimulating discussions of many aspects of the problem.

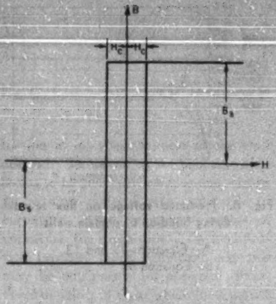


Fig. 1. Rectangular B-H loop for the limiting nonlinear analysis

grain-oriented materials, which exhibit rectangular hysteresis loops, nearly all the domains are known to be aligned parallel or antiparallel with the direction along which a field is normally impressed. On application of such a field, the boundary between oppositely oriented domains (known as a domain or Bloch wall) moves so that a growth of domains directed parallel to the applied field takes place. The motion of a domain wall under the influence of an external field is influenced by several factors.

1. Induced eddy currents reduce the field intensity available at the domain wall.

2. Damping effects other than those caused by eddy currents impede the motion of the wall. This type of damping is known as spin-relaxation damping.

3. The domain walls get caught by imperfections, inclusions, or inhomogeneities in the material, which contribute a frictional force-opposing wall motion, usually known as hysteresis.

4. Inertia effects may influence the acceleration of a domain wall.

5. The wall energy may be a factor influencing wall motion.

The physical principle which determines the formation and motion of Bloch walls is the minimization of the free energy of the system. But even in the idealized case of a single crystal of a ferromagnetic material, its general application is not practicable if a dynamic treatment with changing fluxes is considered. In addition, the nonuniform properties of commercial ferromagnetics on a nonmicroscopic scale makes it practically impossible to express all the forms of free energy mathematically, since they are mostly structure sensitive. For this reason, it is desirable to replace these more rigorous concepts with suitable engineering assumptions.

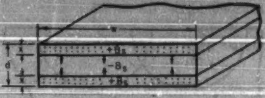


Fig. 2. Flux reversal in the limiting nonlinear analysis

It will be noted that the limiting nonlinear treatment of Wolman and Kaden[4] is completely equivalent to the assumption of an ideally oriented material, with the plane of flux reversal corresponding to a domain wall, though no reference is made to the physical process and only a B-versus-H relationship is postulated. Kittel, Williams, and Shockley[9] presented an analysis of the process of flux reversal in single crystals of silicon iron, based on their knowledge of the domain formations existing in this case. The geometrical arrangement of the test sample of Kittel, et al., differs from that considered here, but this does not detract from the importance of the results of those tests in relation to this study. Kittel, et al., consider first the application of a small magnetic intensity, barely exceeding the coercive force of the material. At such low H, it was found that the domain walls propagate as planar walls, their velocity being proportional to H_e, as defined in the foregoing. The authors show that this result can be predicted if the motion of the walls is assumed to be determined by "classical" eddy currents in the material. This corresponds precisely to the set of assumptions of Wolman and Kaden.[4] The results differ only in so far as the geometry differs. There is close agreement between experiment and theory.

The afore-mentioned limiting nonlinear treatments have been used with good results in the case of many materials and situations.[10] However, they fail in the case of most of the grain-oriented alloys used in magnetic amplifier applications.

A brief review of the literature discussed so far indicates that all the pertinent papers assume basically a planar surface separating oppositely oriented regions and moving from the entire surface of the tape toward its center. The motion is assumed to be determined by a volume distribution of eddy currents such that the value of H at the wall is equal to the coercive force H_c. The philosophy behind the assumption may differ in the papers, from those where even the term domain wall is not used to some of the more recent analyses in which the authors are able to advance very sound reasons in their situation for the existence of such planar domain walls.

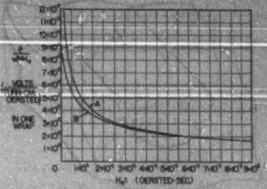


Fig. 3. Predicted induced voltage on flux reversal

A—Equation 1
B—Equation 4

Additional Considerations

The foregoing concepts of flux reversal fail to account for the behavior of grain-oriented nickel-iron alloys and other rectangular hysteresis loop materials under various conditions of flux reversal. Predictions of the induced voltage during flux reversal resulting from the application of a step of mmf F_k, as based on the foregoing concepts, yield the results shown in Fig. 3(A). However, the observed wave forms are as shown in Figs. 14 through 16. Also the effect of tape dimensions on the wave shape and on the numerical value of the time-to-saturation T_s does not agree with the predictions of the analyses quoted.

It is noted that the rate of change of flux (i.e., the switching voltage $e/N=-\dot{\phi}$) is not maximum at the time $t=0$ at which the step of mmf is applied, but reaches such a maximum after a fairly long fraction of the switching process. Recent discoveries in research on ferromagnetism suggest some direct modifications of the limiting nonlinear treatment:

1. The macroscopic or classical eddy currents are not the only damping factor influencing the motion of domain walls.

2. Higher order terms may also enter into the differential equation of motion of a domain wall; specifically, the possibility of an inertia effect in the wall motion should be considered.

3. A critical examination of the concept of a planar wall forming on the entire surfaces of the tape is required.

4. The shape of the moving domain walls may depart from the assumed planar wall.

The points 3 and 4 will form the subject of a detailed analysis in the following section, but points 1 and 2 can be used here in an attempt to improve on the existing treatment.

It is known[11,12] that there are several damping factors in addition to the macroscopic eddy currents considered so far.

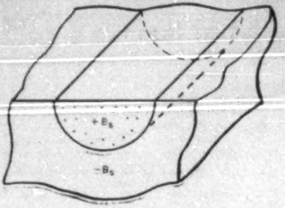


Fig. 4. Shape of an expanding domain

First of all, there are microscopic eddy currents which can be thought of as strictly local eddy currents, e.g., eddy currents of a local nature caused by the uneven motion of the wall around an inclusion or impurity, or as eddy currents taking place over the thickness of the domain wall which has been neglected so far. These currents differ from those already considered in that they depend only on the velocity of the wall (the rate of change of flux, or induced voltage) but not on the position of the wall in the tape since they are of a strictly local nature. There is a second damping factor which appears in the equation of motion of a Bloch wall in an identical manner as the afore-mentioned effect. It is known as spin-relaxation damping and seems to play an important role in the ferrites where eddy currents are practically nonexistent. In its effect spin-relaxation damping cannot be distinguished from the microscopic eddy-current effect in the process of flux reversal considered here.

The equation of motion of a domain wall under the influence of microscopic eddy currents and of the additional effects of microscopic eddy currents and spin-relaxation damping is written

$$\beta_e x\dot{x}+\beta_r\dot{x}=H_e \tag{3}$$

where $\beta_e=0.4\pi 2B_s/\rho$ and β_r is a constant depending both on spin relaxation and microscopic eddy currents.

The switching voltage per turn of pick-up coil becomes

$$\frac{e}{N}=-\dot{\phi}=\frac{-2wH_eB_s}{\sqrt{\frac{\beta_e}{2}H_et+\frac{(\beta_r)^2}{4}}} \tag{4}$$

and switching time

$$T_s=\frac{1}{H_e}\left(\frac{\beta_r}{2}d+\frac{\beta_e}{8}d^2\right) \tag{5}$$

By the inclusion of the term $\beta_r\dot{x}$, the initial switching voltage has been reduced to a finite value, but the wave shapes predicted by the new equation have not changed materially from those obtained previously; see Fig. 3(B).

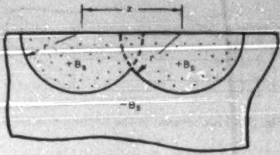


Fig. 5. Domain wall geometry after merging

It has been shown by Becker[13] that domain wall motion is also subject to an effect that appears like an inertia term in the equation of motion. The equation then becomes

$$m\ddot{x}+\beta_r\dot{x}+\beta_e x\dot{x}=H_e \tag{6}$$

The inertia term is thought to be very small, so that $m\ddot{x}$ can be neglected in most lower frequency applications. However, since the value of m could not be ascertained from the literature for the material considered, the possibility of a significant contribution of the inertia term to the process of flux reversal was also investigated. However, an attempt to correlate the foregoing differential equation with the experimentally observed wave forms of switching voltage showed that the equation does not describe the process even to a rough approximation. In the experimental part of this study, additional evidence will be introduced to show that the inertia term is insignificant in the magnetization processes discussed here, at least within the range of speeds of flux reversal considered.

Modified Concepts of Flux Reversal

Since the attempt to explain the process of flux reversal in grain-oriented ferromagnetics, 50-per-cent nickel-iron cores, by means of modification of existing models does not appear successful, a new model will now be formulated, which examines more critically the concept of domain wall formation.

So far it has been assumed that a surface of flux reversal, a domain wall, forms along the lateral sides of the tape and propagates inwards in planar form. Really the formation of regions of reverse magnetization may require considerable energy in an ideally oriented material in which all domain walls have been driven out previously by application of a large field in the opposite direction (i.e., the material is completely saturated). On the other hand, once a domain of reverse magnetization has been created at some point, flux reversal will proceed easily by a growth of this domain.

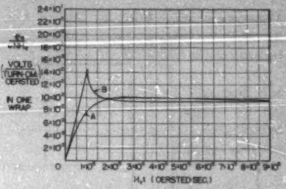


Fig. 6. Predicted voltage on flux reversal during build-up of domain walls

A—Equations 12 and 13
B—Equation 9

In practice, such an ideally saturated condition will not prevail. Judging from the nonperfect squareness of the top of a hysteresis loop it may be expected that upon removal of the saturating field intensity some "nuclei" (though only in a limited number) will be found which are not aligned with the other domains.

Such nuclei probably form around inclusions, impurities, and inhomogeneities (with associated Néel spikes and domains of closure),[14] mostly at the surface of the tape. Therefore, it will be assumed in the analysis that small regions exist near the surface of the tapes which contribute to the process of flux reversal, the tapes being considered having a thickness of 1 to 4 mils. Substantial experimental evidence in favor of such a hypothesis will be introduced later.

It is already clear that for an applied H smaller than the coercive H_c, none of the nuclei of reverse magnetization near or at the surface will grow; most flux changes that take place at such values of H can probably be attributed to departures from the idealized situation envisaged here, and they will be discussed subsequently. As H becomes larger than H_c, some of the nuclei will grow and, thus, cause flux reversal as the Bloch walls sweep over the entire tape. Increases in H are likely to cause more and more of these nuclei to grow. However, when H is large enough, further increases in H can only contribute to the rate of domain wall motion but can no longer add to the number of such walls which are causing flux reversal.

Most of the principles of domain wall motion have already been discussed in some detail and all that remains is to apply them to the situation envisaged here. The shape of the walls can hardly be expected to be planar, even at values of H barely in excess of H_c. For a domain wall with a curvature, the equation of

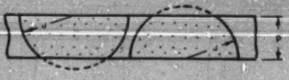


Fig. 7. Domain wall growth at low values of H

motion is modified because of the wall energy. However, this modification seems to be generally a minor one, except for extreme cases of very small radii of curvature. At a very small radius of curvature, the magnitude of the modification becomes more significant and this may be one reason for apparent differences in values of H_c at which the separate nuclei start to grow. In fact, the wall energy may be thought of as a substantial contributary factor to the value of H_c, and this will be discussed in more detail later.

Thus the equation of motion for a planar domain wall will still be retained. It takes the form of equation 3

$$\beta_r v+\beta_e g(x)v=H_e=0.4\pi\frac{NI}{l}-H_c \qquad (7)$$

where v is the velocity measured perpendicularly to the wall at any point. However, when domain wall motion takes place near the surface of the tape, the effect of the macroscopic eddy currents is small compared with that of the microscopic eddy currents and spin-relaxation damping which are independent of the wall position. Therefore, $\beta_e g(x)$ will be neglected altogether, until domain walls have penetrated 1 mil into the tape and have merged to form an essentially continuous wall.

According to some investigators,[15] the nuclei of reverse magnetization may most likely be of an ellipsoidal shape. In case of a thin tape having such regions located at the surface, long, cigar-shaped regions of reverse magnetization will be found intersecting the surface of the tape. Since the regions are long, they will be represented by circular cylinders interesecting the surface at a diameter. Any cross section of the tape will present a similar appearance. Hence, the subsequent motion of the domain walls is treated as a 2-dimensional problem; this approach appears to be a reasonable on-the-average description of the process.

Until adjacent domains merge, it is assumed that the semicircular configuration is preserved for the shape of the expanding domains; see Fig. 4. This assumption, which allows for welcome mathematical simplification, is consistent with the concept of minimization of domain wall energy.[16] No attempt is made to introduce specific assumptions about the initial radii of the semicylindrical nuclei at $t=0$ because of many effects discussed later; in fact, it shall be accepted that the radii are zero at $t=0$.

If a completely random distribution of nuclei in the 2-dimensional problem is assumed, the probability P_1 of finding a nucleus in the distance z is shown to be[17]

$$P_1=1-\epsilon^{-nz}$$

where n is the number of nuclei per centimeter.

The induced voltage per turn caused by two adjacent merging nuclei, Fig. 5, is

$$\frac{e_m}{N}=8B_s r\dot{r}\left(\frac{\pi}{2}-\cos^{-1}\frac{z}{2r}\right),\ 2r\geq z \qquad (8)$$

prior to merging $2r<z$, and $\cos^{-1} z/2r=0$. The expected induced voltage e_T is found by multiplying e_m by the probability density dP_1/dz and integrating over all z. Then the induced voltage per turn during flux reversal caused by one layer of tape is found to be[17]

$$\frac{e_T}{N}=Ax\int_0^{\pi/2}\epsilon^{-x\cos\theta}\,d\theta$$

$$=Ax\frac{\pi}{2}[I_0(x)-L_0(x)] \qquad (9)$$

where

$x=2nr$

$A=4wB_s\dot{r}$

$r\leq d/2$

$r=(H_e t)/\beta_r$

$\dot{r}=H_e/\beta_r$

I_0 = a modified Bessel function of zero order

L_0 = a modified Struve or Weber-Lommel function of zero order with imaginary argument, known as a Nicholson function

Equation 9 is plotted in Fig. 6(B). For tapes thicker than 2 mils, equation 4 applies after the domain walls have penetrated about 1 mil.

When the domain walls growing from opposite surfaces meet near the center of the tape and are annihilated, equation 9 no longer applies. This situation will be considered approximately only because of the mathematical complexities involved. At values of H barely in excess of H_c, there will be so few growing nuclei that most domain walls will not be annihilated till they reach the opposite surface; see Fig. 7. On the other hand, at large values of H, the annihilation of domain walls encountering other walls at the center of the tape may be approximated by placing an imaginary boundary along the center of the tape, Fig. 8, where annihilation of domain walls is assumed to occur at this boundary.

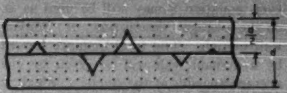


Fig. 8. Domain wall growth at large values of H

The resulting equations for the induced voltage e_{TD} per turn for small H

$$\frac{e_{TD}}{N}=\frac{e_T}{N}-Ax\cos^{-1}\frac{2dn}{x} \qquad (10)$$

for larger H

$$\frac{e_{TD}}{N}=\frac{e_T}{N}-Ax\cos^{-1}\frac{dn}{x} \qquad (11)$$

The previous analysis has been based on the assumption of a completely random distribution of growing nuclei. It may be of interest to carry out a similar analysis with the alternate assumption of an ordered distribution, i.e., the growing nuclei are assumed to be evenly spaced.

For this new case, the switching voltage prior to merging takes the value

$$\frac{e_T'}{N}=4B_s\pi r\dot{r}nw=\frac{\pi}{2}Ax \qquad (12)$$

where

$$2r\leq s=\frac{1}{n}$$

and the prime is used to denote the new case considered. After merging

$$2r>s$$

$$\frac{e_T'}{N}=4\pi rB_s w\sin^{-1}\frac{s}{2r}(2nr)\dot{r}=Ax\sin^{-1}\frac{1}{x} \qquad (13)$$

Here, e_T'/NA takes a maximum value of $\pi/2$ compared with that of about 1.096 for the random distribution of growing nuclei; see Fig. 6(A).

e_{TD}' is obtained from e_T' in the same manner as e_{TD} has been obtained from e_T in equations 10, and 11, i.e., by subtraction of a term $Ax\cos^{-1}2dn/x$, or ($Ax\cos^{-1}dn/x$). It will be observed that the expressions derived for e_T, e_{TD}, e_T', and e_{TD}' represent the domain wall area per unit length of tape in each case, as a function of time, if n and H_e and thus $\dot{r}$ are considered to be constants.

Influence of Core Dimension and H on T.

The time required for flux reversal or the switching time T_s will be considered for different ranges of values of H and core dimensions. At small values of H_e, very few nuclei will grow, and the wall motion will be quite nonuniform. How-linear capacitor. Here the peak stored linear capacitor is the average of the two half cycle cycle values. Of course, using the definitions for the loss factor D of a linear capacitor in terms of an equivalent series or parallel resistance, $D=R_s\omega C$ or $D=1/(R_p\omega C)$ respectively, the equivalent series or parallel resistances for the nonlinear capacitor can be calculated.

The equivalent capacitance and dissipation factor of the nonlinear capacitor were calculated for each loop. Figs. 9, 10, and 11 show these results. Fig. 9 displays the variation of the equivalent capacitance of the nonlinear capacitor with bias for various values of peak positive alternating voltage. For large values of alternating voltage, these curves show that the capacitance becomes nearly independent of bias. Thus, it is to be expected that the amplification of dielectric amplifier circuits in which the alternating voltage across the nonlinear capacitor is large will be less than that obtained with a smaller alternating voltage. These curves also show that, as the bias increases, the capacitance tends to become independent of the alternating voltage.

Fig. 10 demonstrates the latter point more clearly. Here equivalent capacitance is plotted against the peak positive alternating voltage for several values of bias voltage. For small values of bias voltage, the capacitance varies considerably with alternating voltage, but becomes nearly independent of alternating voltage for large bias voltages. Nonlinear resonance phenomena[7,8] will be less noticeable at high values of bias voltage than at low bias voltages.

In Fig. 11, the dissipation factor is plotted versus bias voltage for various peak positive alternating voltages. All of these curves show a trend of decreasing losses as the bias voltage is increased. This was evident upon examining the hysteresis loops, as the area of the loops decreased noticeably with increasing bias voltage, Figs. 7(A) and (B). These curves show that, for large values of bias higher values of resonant circuit Q are possible, which would tend to increase the amplification of a resonant dielectric amplifier.[1] Also, the temperature rise of the nonlinear capacitor will be less for a given energy input at high bias voltages.

Conclusion

The preceding representation of the nonlinear capacitor by equivalent linear elements simplifies analysis of the dielectric amplifier by making possible the application of linear circuit theory to circuits where the alternating voltage drop across the nonlinear capacitor is not small. Graphical methods can be applied to give solutions of the large signal properties and, by expressing the curves describing the behavior of the equivalent linear elements in terms of their partial derivatives about the quiescent point, the small signal properties can be analyzed in the same manner as vacuum-tube or transistor circuitry has been analyzed. Derivation of an equivalent circuit for the dielectric amplifier, determination of stability conditions, and optimization of circuit behavior are problems that are considerably reduced by this linear representation. Furthermore, the accurate prediction of resonant circuits where the Q is controlled by the nonlinear capacitor is now possible.

References

1. DIELECTRIC AMPLIFIERS, **G. W. Penney, J. R. Horsch, E. A. Sack.** *AIEE Transactions*, vol. 72, pt. I, Mar. 1953, pp. 68–79.

2. FERROELECTRICS AND THE DIELECTRIC AMPLIFIER, **W. P. Mason, R. F. Wick.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, Nov. 1954, pp. 1606–20.

3. BARIUM TITANATE AND BARIUM STRONTIUM TITANATE AS NON-LINEAR DIELECTRICS, **S. Roberts.** *Doctoral Dissertation*, Massachusetts Institute of Technology, Cambridge, Mass., 1946.

4. THE EFFECT OF MINOR CONSTITUENTS IN HIGH DIELECTRIC CONSTANT TITANATE CAPACITORS, **W. W. Coffeen.** *AIEE Transactions*, vol. 72, pt. I, 1953, pp. 704–09.

5. POWER GAIN OF A DIELECTRIC AMPLIFIER, **S. Zisk.** *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 10, 1954, pp. 90–95.

6. AN INVESTIGATION OF THE RESONANT CIRCUIT DIELECTRIC AMPLIFIER, **E. A. Sack.** *Doctoral Dissertation*, Carnegie Institute of Technology, Pittsburgh, Pa., 1954.

7. FREQUENCY RESPONSE OF A RESONANT DIELECTRIC AMPLIFIER, **G. W. Penney, E. A. Sack, E. R. Wingrove.** *AIEE Transactions*, vol. 73, pt. I, May 1954, pp. 119–24.

8. DIELECTRIC AMPLIFIER ANALYSIS, **J. R. Horsch.** *Doctoral Dissertation*, Carnegie Institute of Technology, Pittsburgh, Pa., 1953.

9. A MATHEMATICAL ANALYSIS OF A DIELECTRIC AMPLIFIER, **L. A. Pipes.** *Journal of Applied Physics*, New York, N. Y., vol. 23, no. 8, Aug. 1952, pp. 818–24.

10. APPLICATIONS OF INTEGRAL EQUATIONS TO THE SOLUTION OF NONLINEAR ELECTRIC CIRCUIT PROBLEMS, **L. A. Pipes.** *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 445–50.

11. MATHEMATICAL ANALYSIS OF A SERIES CIRCUIT CONTAINING A NONLINEAR CAPACITOR, **Louis A. Pipes.** *Ibid.*, Jul. 1953, pp. 238–44.

12. A FERROELECTRIC AMPLIFIER, **H. Urkowitz.** *Journal*, Franklin Institute, Philadelphia, Pa., vol. 254, Dec. 1952, pp. 517–30.

13. ANALYSIS OF BASIC DIELECTRIC AMPLIFIER CIRCUIT, **Shou-Hsien Chow.** *Journal of Applied Physics*, New York, N. Y., vol. 25, Oct. 1954, pp. 1297–1301.

◆

Discussion

E. A. Sack and **P. N. Wolfe** (Westinghouse Research Laboratories, Pittsburgh, Pa.): The authors of this paper have pointed out that the losses in the ferroelectric ceramic raised the temperature of their experimental specimen above the temperature of the silicone fluid. However, no information has been given as to the actual temperature of the ceramic at each point on the experimental curves. Without this information, both the data which have been presented and the conclusions the authors have drawn from them are strictly applicable only to the particular experimental configuration which was used.

At least two of the conclusions drawn by the authors are invalid when the hysteresis loops are studied at a constant ceramic temperature. First, when the ceramic temperature is maintained at a value slightly above the apparent Curie temperature, the hysteresis loops for a high-K ferroelectric are not accurately represented by straight lines. Second, under the same conditions, the major and minor loops actually do tend to nest, and the so-called mean polarization curve is a very reasonable approximation to the material behavior for use in computations where a single-valued representation is required.

The major hysteresis loops of Fig. 12 are for a high-K barium strontium titanate ceramic capacitor obtained from the same manufacturer as the material used by the authors. The peak of the applied alternating potential on the capacitor approached the maximum permitted by breakdown of the specimen. For the loops of Fig. 12(A), the specimen was suspended in a temperature-regulated silicone-fluid bath by a pair of thin wire leads. The more vertical loop was photographed immediately after the test voltage had been applied to the sample and the initial electric transient had subsided. Another exposure was made about 30 seconds later, at which time the loop had fallen over as shown. For Fig. 12(B), the sample was placed in an electrode structure which was specially designed to carry away the heat generated in the ceramic. This structure was immersed in the same silicone-fluid bath. Again, two exposures were made. This time, however, it can be seen that there was no significant change in the shape of the loop.

The results of the preceding test are in agreement with the conclusions of the authors in that the loops appear to fall over when the electrode structure is not adequate to maintain the ceramic temperature at the bath temperature. This test also shows that the special electrode structure used by us was adequate at the frequency used in the following tests.

Fig. 13 shows a nest of major hysteresis loops obtained at constant ceramic temperature. It can be seen that the smaller loops do not fall outside of the larger loops.

Fig. 12. Double exposures which show the effect of internal heating on the hysteresis loop of a ferroelectric ceramic

A—Inadequate cooling
B—Adequate cooling

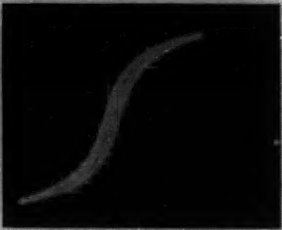


Fig. 13. Nest of major hysteresis loops for the ferroelectric ceramic

It should also be noted that the locus of the tips of the loops appear to be a far better approximation to the electrical behavior of the capacitor than a straight line.

Fig. 14 (A) is a minor loop for the ferroelectric ceramic capacitor. The direct bias on the ceramic for this loop was 300 volts and the alternating potential swing was 580 volts, zero to peak. In Fig. 14(B) this same loop has been photographed through the (photographic) negative of a major loop formed by an alternating voltage which reached the same negative peak value. It is apparent that the minor loop lies within the major loop and does not, as the authors contend, fall far outside of this boundary.

Fig. 15 is a minor loop for the same bias, but with a 427-volt zero-to-peak alternating-potential swing applied to the capacitor. Again, this loop lies within the major loop. Similar photographs have been made for a wide range of bias and alternating voltages. In no case do the minor loops deviate significantly from the path of the major loop locus.

Finally, Fig. 16(A) shows the area covered by a very large number of minor loops. This figure is compared with the major dynamic loop for the same maximum potential swing. The congruence of the two figures is readily apparent.

The preceding tests emphasize the importance of either maintaining or measuring the actual ceramic temperature when taking data concerning the hysteresis loops of the ferroelectric dielectric. These tests also indicate that, when the ceramic temperature is constant, the mean polarization curve is a more accurate single-valued approximation to the electrical behavior of the material than the straight lines proposed by the authors. Finally, the tests show that it is highly desirable to provide an adequate cooling structure for the ferroelectric capacitors in a dielectric amplifier, since the nonlinearity and, therefore, the control capabilities of the ceramic are diminished by severe internal heating.

Fig. 14. Minor loop for an alternating potential of 580 volts zero to peak and a direct bias of 300 volts

A—Minor loop
B—Same loop photographed through the (photographic) negative of a major loop

In closing, it should be stressed that we do not claim that the data and conclusions presented by the authors are incorrect for the specific experimental configuration which they have examined. We do contend, however, that since the authors did not know the temperature of the ceramic in their tests, their conclusions must not be interpreted as generalizations and are valid only for one particular thermal geometry.

Fig. 15. Minor loop for an alternating potential of 427 volts zero to peak and a direct bias of 300 volts

A—Minor loop
B—Same loop photographed through the (photographic) negative of a major loop

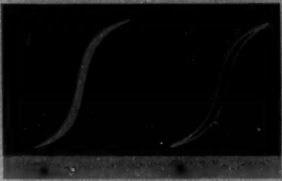


Fig. 16. Comparison of the regions on the Q, V plane which are occupied by a major loop and by a set of minor loops

A—Minor loops for various bias potentials
B—Major loop

Earl Wingrove, Louis Depian, and **W. Lee Shevel:** We wish to thank Mr. Sack and Mr. Wolfe for their discussion, but find that the conditions used by them are not those ordinarily encountered in practical dielectric amplifiers. The object of the paper is to show a representation of the nonlinear capacitor when used in typical dielectric amplifier circuitry. There is little use to discuss the capacitor behavior in conditions foreign to such circuitry. The goal of the amplifier is to provide a device which has certain desirable characteristics, such as reasonably high-voltage gain, acceptable bandwidth, good stability, and reasonable operating temperature. The bandwidth depends upon the carrier frequency used. It has been shown (see ref. 7 of the paper) that, with carriers of the order of 1 megacycle, a frequency response flat to 10 kc is possible. We used a carrier of about 0.6 megacycle at which the bandwidth was 4 to 5 kc. Operation of a dielectric amplifier at carrier frequencies below, say 15 kc, may not be considered too practical. From a temperature point of view, a convenient operation would be with an ambient of room temperature. We used an ambient of 30 degrees centigrade. The discussers, using the same material, maintained the temperature slightly above the Curie temperature (12 degrees centigrade). Examining the loops of this paper and those of the discussers, it can be seen that the loops at 30 degrees centigrade are very linear compared to the loops at about the Curie temperature. It is apparent that the discussers' objection to the linear representation is valid only in that particular case of operation near the Curie temperature.

Mr. Sack and Mr. Wolfe do not mention the frequency at which their curves were obtained. This is unfortunate since we found that an adequate electrode structure is indeed easily obtainable at low carrier frequencies. This, of course, will allow the nesting of the hysteresis loops, as shown by the discussers and in the paper; see Fig. 4. However, at more useful higher frequencies of 0.5 megacycle, such an electrode structure is not too practical and if then the temperature of the bath is kept constant, the loops will fall over as shown in the paper. Furthermore, under normal operating conditions, it can be shown that because of the low heat conductivity of the material, the temperature of the center of the ceramic may be higher (10 degrees centigrade is not uncommon) than that of the silvered surface. Therefore, the temperature of the ceramic is not necessarily constant even if this is accomplished for the silvered surface through large electrodes.

Another point is also of interest. Using a singled-valued representation such as the mean polarization curve, as the discussors suggest, and representing the losses by some constant independent of bias or peak alternating voltage, there is no way of taking into account the variation of losses of the capacitor. With the loop interpretation as given in the paper, this is done and the losses are found as a function of bias and peak alternating voltage. The dielectric amplifier under the operating conditions stated in the paper yielded a stable voltage gain of 15 which is significantly higher than that reported in the literature to date.

Power Transistor Switching Circuit

CHAANG HUANG
NONMEMBER AIEE

EDWIN SLOBODZINSKI
NONMEMBER AIEE

TO STUDY power transistor switching circuits, it is necessary to measure relevant d-c transistor parameters as functions of current and temperature. Based on the measured results, design methods can be developed for high-power bistable switching circuits which have good trigger sensitivity and d-c stability over wide temperature ranges.

Among the d-c transistor parameters, the most temperature-dependent parameter is the reverse collector current. If the operating voltage is quite small compared to the breakdown voltage, the reverse collector current is composed of two parts:

1. A leakage current I_{cl} which varies linearly with the applied voltage, and which is relatively independent of temperature.
2. A temperature-sensitive but voltage-independent component I_{ct}.

When the ambient temperature is raised, the temperature-sensitive component of the reverse collector current will increase in an exponential fashion. This current flows in the base region of the transistor and produces a transversal voltage drop which tends to bias the center of the emitter forward.[1] When a certain temperature is reached, the emitter will start to inject minority carriers and the collector current will be increased, resulting in an increase in power dissipation and junction temperature. The increase in junction temperature will further increase the reverse collector current and, consequently, increase the transversal voltage drop and emitter injection. This is a cumula-effect. It causes an OFF switching transistor to turn ON without the application of a triggering pulse. A means of compensating for this instability is to use a suitable base bias to prevent the emitter from injecting carriers. An excessive base bias, on the other hand, will not allow the switching transistor to have a stable ON state.

It is the purpose of this paper to describe the methods and results of the measurements of the d-c parameters of a *2N68* power transistor. With these data, several power transistor switching circuits are considered with an emphasis on the design of the base bias for temperature considerations. These design techniques can be easily adopted to permit their application to most direct-coupled switching circuits with the use of p-n-p or n-p-n transistors.

Measurement of Parameters

The d-c equivalent circuit of a transistor is shown in Fig. 1.

Where

r_e, r_b, r_c = d-c emitter, base, and collector resistances respectively
I_{ct} = temperature-sensitive component of the reverse collector current
α = the large signal or d-c amplification factor

α is defined as[2]

$$\alpha = \frac{I_c(V_c, V_e) - I_c(V_c, 0)}{I_e(V_c, V_e)} \qquad (1)$$

The measurement of r_e is made by passing a forward current through the emitter base circuit with the collector terminal open-circuited. If a high impedance voltmeter is connected across the collector emitter terminals, the voltage read is essentially $I_e r_e$. With I_e measured, r_e is determined as a function of the forward current.

The values of r_e versus I_e of a *2N68* power transistor are plotted in Fig. 2. The emitter resistance takes a high value when the emitter current is in the reverse direction. This value is of minor importance in the design of power transistor circuits.

Similarly, if a forward current is passed through the emitter base circuit, the voltage read across the collector base terminals determines r_b as a function of forward base current, as shown in Fig. 3.

When the emitter is biased in the reverse direction, the emitter current is near to zero. The voltage distribution caused by the flow of the reverse base current appears between the emitter and base terminals. This voltage is dependent on the magnitude of the reverse base current. The effect of this voltage on circuit performance can be lumped into an apparent base resistance r_b, defined as the ratio of the emitter-base voltage to the reverse base current. The variations of r_b versus the reverse base current for different values of ambient temperature are plotted in Fig. 4.

To measure the r_c, the reverse collector characteristic is first obtained, as shown in Fig. 5. The reverse collector current can be written as

$$I_{c0} = I_{ct} + \frac{V_c}{r_c}$$

Thus

$$r_c = \frac{dV_c}{dI_{c0}}$$

A plot of r_c versus V_c is given in Fig. 6. The variation in I_{c0} with junction temperature at a given collector bias is illustrated in Fig. 7.

The large-signal current-amplification factor α can be measured with either d-c or pulse methods. The collector is placed at a specified voltage while the emitter current is increased from zero to a certain value I_e. The ratio of the change in the collector current from I_{c0} to the value of I_e is, by definition, the d-c amplification factor at emitter current I_e and collector voltage V_c. Fig. 8 shows the variation of α versus I_e.

Power Transistor Flip-Flop Circuit Design

A conventional p-n-p junction transistor flip-flop circuit is shown in Fig. 9 with a compensating base bias indicated by R_b and E_{bb}. First consider the OFF transistor (transistor *T-1* in Fig. 9). Temperature compensation is required to keep this transistor OFF. Fig. 10 is a simplified diagram of the OFF transistor circuit, where the current generator

$$\frac{E_{cc}}{R_{1c} + R_{2f} + r_{2b}}$$

represents the current fed to the base of the ON transistor. $E_{c,ON}$ is the ON transistor's collector voltage (the saturation collector voltage of the ON transistor no. 2). To keep transistor *T-1* in the OFF state, it is required that

$$I_{1e} < 0$$

or

$$I_{1b} \leq I_{1c}$$

Circuit analysis, Appendix I, leads to the requirement that

$$R_{1b} \leq \frac{E_{bb}R_{1f} - r_{1b}(T)I_{c0}(T)R_{1f}}{I_{c0}(T)(r_{1b}(T) + R_{1f}) + E_{c,ON}} \qquad (2)$$

where

T = temperature
I_{c0} and r_b = the temperature-dependent parameters

Paper 56-251, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing January 16, 1956.

CHAANG HUANG is with Sylvania Electric Products, Inc., Woburn, Mass., and EDWIN SLOBODZINSKI is with the International Business Machines Corporation, Poughkeepsie, N. Y.

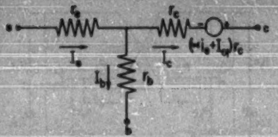


Fig. 1. Transistor d-c equivalent circuit

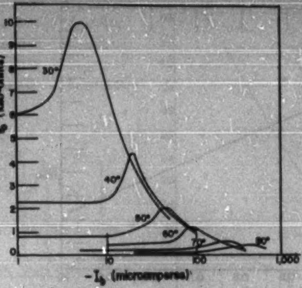


Fig. 4 (above). r_b versus I_b

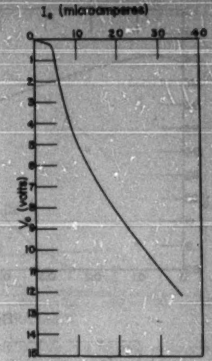


Fig. 5 (right). V_c—I_e characteristic

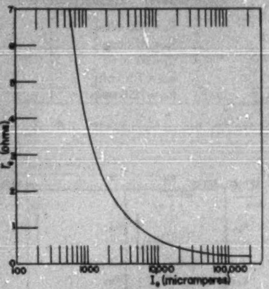


Fig. 2. Forward r_e versus I_e

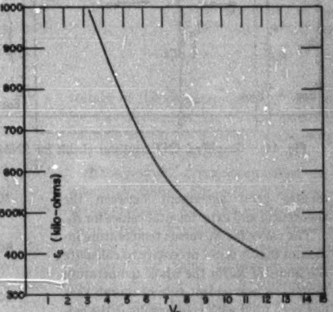


Fig. 6 (right). r_c versus V_c

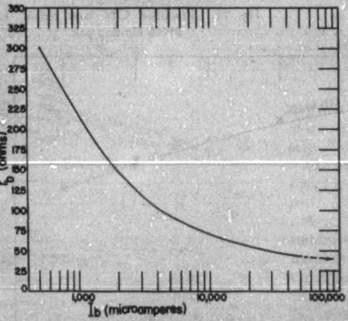


Fig. 3. Reverse r_b versus I_b

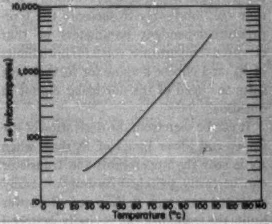


Fig. 7. I_{co} versus temperature at 12 volts

which thus places an upper limit on R_b.

Next consider the requirements for transistor *T-1* to be in the ON state and obtain a lower limit for R_{1b}. A simplified ON transistor circuit is shown in Fig. 11. This circuit was drawn, assuming that the ON i_e r_e drop is very small compared with the i_b r_b drop.

For a transistor to be in the ON state, its collector current has to be saturated

$$I_{1c}=\frac{E_{cc}}{R_{1c}}\leq\frac{\alpha_1}{1-\alpha_1}I_{1b} \tag{3}$$

where α_1 is the d-c amplification factor.

Circuit analysis, Appendix I, leads to lower limit for R_{1b}

$$R_{1b}\geqq\frac{(R_{1f}+R_{1c})[\beta_1 R_{1c}E_{bb}+r_{1b}(T)E_c]}{E_c[\beta_1 R_{1c}+R_{1f}+R_{2c}+r_{1b}(T)]} \tag{4}$$

where

$$\beta_1=\frac{\alpha_1}{1-\alpha_1}$$

Similarly

$$\frac{(R_{2f}+R_{2c})[\beta_2 R_{2c}E_{bb}+r_{2b}(T)E_{cc}]}{E_{cc}[\beta_2 R_{2c}+R_{2f}+R_{2c}+r_{2b}(T)]}\leq R_{2b}\leq \frac{E_{bb}R_{2f}-R_{2f}r_{2b}(T)I_{c0}(T)}{I_{c0}(T)(r_{2b}(T)+R_{2f})+E_{c,\text{ON}}}$$

To illustrate the design procedure, a sample symmetrical flip-flop circuit is designed. If $E_{cc}=12$ volts, $E_{bb}=7.5$ volts and $I_{c,\text{ON}}\doteq 1.2$ amperes, then $R_c=10$ ohms, with $\beta\doteq 10$, and $i_b\doteq 120$ milliamperes.

The consideration for R_f is that it should provide sufficient feedback current while not causing an excessive drop in the OFF transistors collector load; i.e.

$$R_e\leq R_f\leq\frac{E_c}{I_b}-(R_c+r_b)$$

Here the values of R_f are 50 ohms.

Substituting these values and the temperature-dependent parameters of the transistor into equations 2 and 4, the upper and lower limits of permissible R_b are obtained, which are plotted and compared with the experimental upper limit of R_b in Fig. 12. The comparison in-

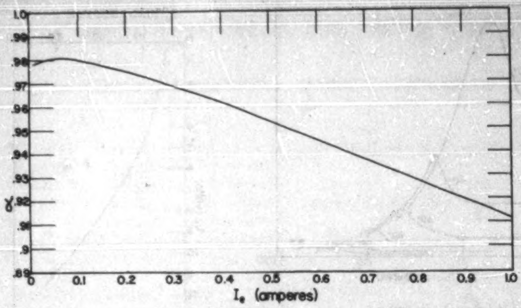


Fig. 8. α versus I_e

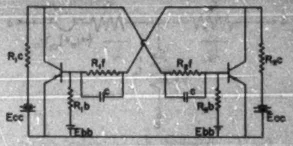


Fig. 9 (above). 2N68 flip-flop pair

$R_b = 300$ ohms
$R_e = 10$ ohms
$R_f = 50$ ohms
$E_{bb} = 1.5$ volts
$E_{cc} = 12.0$ volts

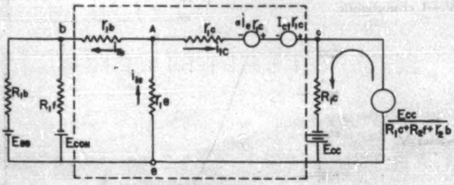


Fig. 10. Simplified OFF transistor circuit for 2N68 flip-flop pair

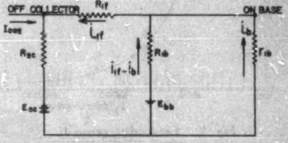


Fig. 11. Simplified ON base circuit for 2N68 flip-flop pair

dicates good agreement between the theoretical and experimental values for R_b.

The curve for R_b versus temperature indicates that it is not necessary to calculate the limits of R_b for the whole temperature range. A convenient rule of thumb that has been used with good results is the calculation of R_b for the maximum junction temperature which is to be realized. It is now only necessary to measure the temperature-dependent parameters at that elevated temperature, solve for both limits of R_b, and choose a value for R_b that best suits the need of the particular circuit in question.

It can be seen from the derivation of the upper and lower limits of R_b that, when R_b is near the upper limit the transistor can be easily switched ON but not easily switched OFF, and, if R_b is near the lower limit, the transistor can be easily switched OFF but not easily switched ON. For optimum trigger sensitivity, sometimes it is favorable to use a temperature-sensitive element,[3] the resistance of which lies between the two limits.

It is to be noted that, in the expression for the lower limit of R_b, the d-c amplification factor β is actually temperature-dependent. But the variation is found to be small. For a conservative design, a value of R_b should be used which is higher than the lower limit of R_b that was calculated with the assumption that α was constant.

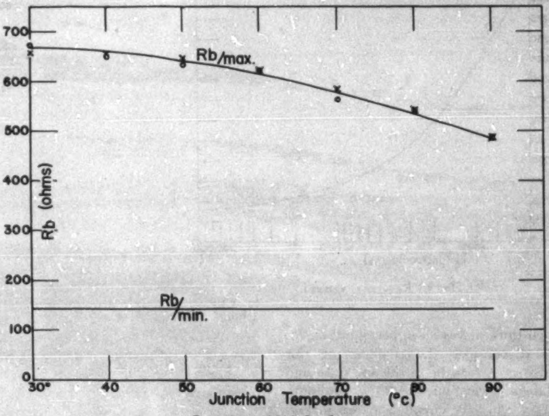


Fig. 12. 2N68 flip-flop

Upper and lower limits of R_b versus junction temperature: o—Calculated, x—Experimental

2N34 Flip-Flop Circuit with Cascaded Power Transistors

The power transistor circuit described in the foregoing was found to possess relatively poor trigger sensitivity. In other words, relatively large triggering power is required. Fig. 13 shows a circuit with a much improved trigger-sensitivity characteristic. This circuit consists of two *2N34* transistors in a flip-flop circuit driving two cascaded *2N68* transistors, resulting in an output current of 1 ampere.

Following the same temperature considerations, $i_e<0$ for the OFF transistors and $I_b>E_c/(R_c\beta)$ for the ON transistors. A circuit analysis, Appendix II, yields criteria for the choice of base bias resistances, which are listed in the following equations.

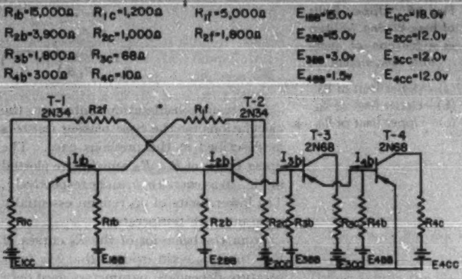

R1b=15,000Ω
R2b=3,900Ω
R3b=1,800Ω
R4b=300Ω
R1C=1,200Ω
R2C=1,000Ω
R3C=68Ω
R4C=10Ω
R1f=5,000Ω
R2f=1,800Ω
E1BB=15.0v
E2BB=15.0v
E3BB=3.0v
E4BB=1.5v
E1CC=18.0v
E2CC=12.0v
E3CC=12.0v
E4CC=12.0v
T-1
2N34
R2f
R1f
T-2
2N34
T-3
2N68
T-4
2N68
I1b
I2b
I3b
I4b
R1C
R1b
R2b
R2C
R3b
R3C
R4b
R4C
E1CC
E1BB
E2BB
E2CC
E3BB
E3CC
E4BB
E4CC

Fig. 13. 2N34 flip-flop pair with two 2N68 transistors cascaded

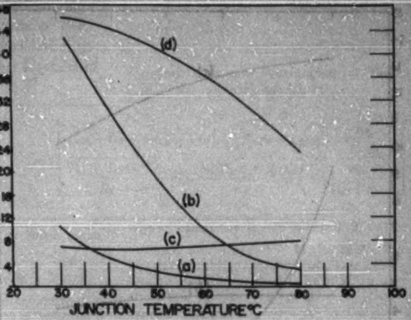


Fig. 14. Upper limits of base-biasing resistors as a function of junction temperature

(a)—Upper limit of R_{4b}

(b)—Upper limit of R_{3b}

(c)—Upper limit of R_{2b}

(d)—Upper limit of R_{1b}

$$\frac{E_{1bb}R_{1f}-I_{1c0}r_{1b}R_{1f}}{E_{2c,ON}+I_{1c0}R_{1f}+I_{1c0}r_{1b}}>R_{1b}>\frac{\left(E_{1bb}+\frac{E_{1cc}r_{1b}}{R_{1c}\beta_1}\right)(R_{1f}+R_{2c})}{E_{2cc}-\frac{E_{1cc}}{R_{1c}\beta_1}(R_{1f}+R_{2c}+r_{1b})}$$

$$\frac{R_{2f}(E_{2bb}-E_{3bb}+I_{3c0}R_{3b}-I_{2c0}r_{2b})}{E_{1c,ON}+E_{3bb}+I_{2c0}r_{2b}-I_{3c0}R_{3b}+I_{2c0}R_{2f}}>R_{2b}>$$

$$\frac{E_{2bb}+I_{4b}r_{4b}+I_{3b}r_{3b}+(E_{2cc}-I_{4b}r_{4b}-I_{3b}r_{3b})\,\frac{r_{2b}}{R_{2c}\beta_2}}{\frac{E_{1cc}-I_{4b}r_{4b}-I_{3b}r_{3b}}{R_{2f}+R_{1c}}-\frac{E_{2cc}-I_{4b}r_{4b}-I_{3b}r_{3b}}{R_{2c}\beta_2}\left(1+\frac{r_{2b}}{R_{2f}+R_{1c}}\right)} \qquad (5)$$

$$\frac{E_{3bb}-E_{4bb}}{I_{3c0}}+R_{4b}-r_{3b}>R_{3b}>\frac{E_{3cc}+I_{4b}r_{4b}+\left(\frac{E_{3cc}-I_{4b}r_{4b}}{R'_{c}\beta_3}\right)r_{3b}}{\frac{E_{3bb}-I_{4b}r_{4b}}{R_{3c}}-\left(1+\frac{r_{3b}}{R_{3c}}\right)\left(\frac{E_{3cc}-I_{4b}r_{4b}}{R_{3c}\beta_3}\right)}$$

$$\frac{E_{4bb}}{I_{4c0}(T+\Delta T)}-r_{4b}(T+\Delta T)>R_{4b}>\frac{\frac{E_{4cc}r_{4b}}{R_{4c}\beta_4}+E_{4bb}}{\frac{E_{3cc}}{R_{3c}}-\frac{E_{4cc}}{R_{4c}\beta_4}\left(1+\frac{r_{4b}}{R_{3c}}\right)}$$

where

$$I_{4b}=\frac{\frac{E_{3cc}}{R_{3c}}-\frac{E_{4bb}}{R_{4b}}}{1+\frac{r_{4b}}{R_{4b}}+\frac{r_{4b}}{R_{3c}}}$$

and

$$I_{3b}=\frac{\frac{E_{2cc}-I_{4b}r_{4b}}{R_{2c}}-\frac{E_{3bb}+I_{4b}r_{4b}}{R_{3b}}}{1+\frac{r_{3b}}{R_{2b}}+\frac{r_{3b}}{R_{3b}}}$$

To insure d-c stability for an ambient temperature range of 25 to 80 degrees centigrade, the circuit parameters were chosen so that the upper limits of the base load resistance would not intersect their corresponding lower limit in the desired temperature range. In the design of a circuit of this type, the detailed considerations to be made are as follows:

The circuit parameters of *T-4* are designed to produce the desired output current. The values of R_{3c} and R_{3c} are to be chosen while keeping in mind that emitters of *T-3* and *T-2* are at finite negative voltages when the transistors are in the ON state, these negative voltages being the successive $I_b r_b$ drops. The choice of E_{3bb} and E_{2bb} are not as random as is the choice of E_{4bb}. They must be chosen to compensate for the positive emitter voltage of the OFF transistors as well as for the increase of I_{c0}. These positive voltages appear as a result of the successive positive base biases.

In the case of *T-2* there is an added consideration; namely, the feedback loop consisting of R_{2f} and R_{1c}. The upper limit of R_{2b} is strongly dependent on R_{2f} whereas R_{1c} has a minimum value dependent on the maximum value allowable on collector current for *T-1*. A compromise must be made providing sufficient feedback while still limiting the ON collector current of *T-1*.

With the properly chosen parameters in the circuit of Fig. 13 for an output of 1 ampere and 12 volts, the upper limits of R_{4b}, R_{3b}, R_{2b}, and R_{1b} are plotted in Fig. 14 as curves *a*, *b*, *c*, and *d*, respectively. The minimum values of R_{4b}, R_{3b}, R_{2b}, and R_{1b} are 50 ohms, 1,000 ohms, 3,600 ohms, and

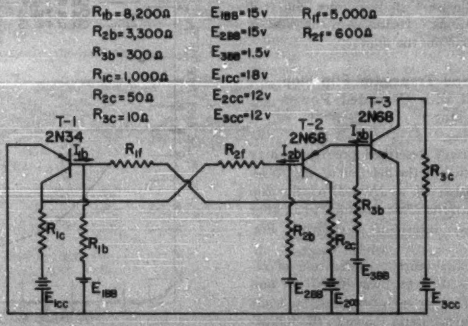

R1b=8,200Ω
R2b=3,300Ω
R3b=300Ω
R1C=1,000Ω
R2C=50Ω
R3C=10Ω
E1BB=15v
E2BB=15v
E3BB=1.5v
E1CC=18v
E2CC=12v
E3CC=12v
R1f=5,000Ω
R2f=600Ω
T-1
2N34
T-2
2N68
T-3
2N68
I1b
I2b
I3b
R1f
R2f
R1C
R1b
R2b
R2C
R3b
R3C
E1CC
E1BB
E2BB
E2C
E3BB
E3CC

Fig. 15. 2N34–2N68 pair with cascade circuit for high output currents

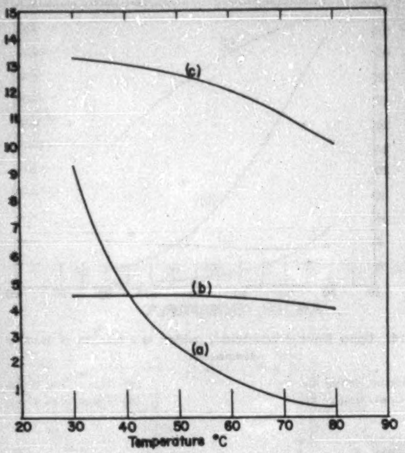


Fig. 16. Upper limits of base-biasing resistors as a function of junction temperature

(a)—Upper limit of R_{3b}
(b)—Upper limit of R_{2b}
(c)—Upper limit of R_{1b}

13,480 ohms respectively. In the foregoing calculation the final value of the base-biasing resistors were chosen so that the trigger sensitivity would be good while a change in transistor parameters would not result in unstable operation.

With the requirement of R_b versus temperature fairly well understood, the entire design procedure may be simplified. Since the upper limits of R_b, R_{2b}, and R_{3b} decrease with temperature, it is simply a matter of measuring the relevant temperature-dependent parameters at the maximum required temperature of operation, solving for the limits of R_b at that temperature, and making a choice of values to suit the needs of the circuit in question. In the case of R_{3b} it may be necessary to plot the curve of the upper limit so that unstable operation at low temperatures may be avoided. In all cases, each particular application may require different considerations. However, the basic method remains the same.

2N34-2N68 Transistor Flip-Flop Circuit with Cascaded Power Transistor

The circuit described previously may be simplified by the deletion of one *2N34*. This consists of a *2N34-2N68* flip-flop circuit with the *2N68* cascaded to another *2N68* power transistor, as shown in Fig. 15.

The temperature consideration of requiring $I_c<0$ for each OFF transistor and $I_b>E_{cc}/(R_c\beta)$ for each ON transistor leads to the design criteria shown in the following; see Appendix III

$$\frac{E_{1bb}R_{1f}-I_{1co}r_{1b}R_{1f}}{E_{2c,\mathrm{ON}}+I_{1co}R_{1f}+r_{1b}I_{1co}}>R_{1b}>$$

$$\frac{\left(E_{1bb}+\dfrac{E_{1cc}r_{1b}}{R_{1c}\beta_1}\right)(R_{1f}+R_{2c})}{E_{2cc}-\dfrac{E_{1cc}}{R_{1c}\beta_1}(R_{1f}+R_{2c}+r_{1b})}$$

$$\frac{R_{2f}(E_{2bb}-E_{3bb}+I_{3co}R_{3b}-I_{2co}r_{2b})}{E_{1c,\mathrm{ON}}+E_{3bb}+I_{2co}r_{2b}-I_{3co}R_{b3}+I_{2co}R_{2f}}>R_{2b}>$$

$$\frac{E_{2bb}+I_{3b}r_{3b}+(E_{2cc}-I_{3b}r_{3b})\dfrac{r_{2b}}{R_{2c}\beta_2}}{\dfrac{E_{1cc}-I_{3b}r_{3b}}{R_{2f}+R_{1c}}-\dfrac{E_{2cc}-I_{3b}r_{3b}}{R_{2c}\beta_2}\left(1+\dfrac{r_{2b}}{R_{2f}+R_{1c}}\right)} \tag{6}$$

$$\frac{E_{3bb}}{I_{3co}(T+\Delta T)}-r_{3b}(T+\Delta T)>R_{3b}>$$

$$\frac{\dfrac{E_{3cc}r_{3b}}{R_{3c}\beta_3}+E_{3bb}}{\dfrac{E_{2cc}}{R_{2c}}-\dfrac{E_{3cc}}{R_{3c}\beta_3}\left(1+\dfrac{r_{3b}}{R_{3c}}\right)}$$

The same considerations enter into the calculations for the base-biasing resistors as described in the previous case. The upper limits of R_{3b}, R_{2b} and R_{1b} are plotted in Fig. 16 as curves *a*, *b*, and *c* respectively. The lower limits of R_b remain essentially constant with temperature.

From the behavior of the R_b curves of Fig. 16, it is again evident that the temperature-dependent parameters need only to be measured at the maximum operating temperature required. Solving for the upper and lower limits of R_b at this temperature provides a range of values of permissible R_b. The choice of R_b depends on the particular circuit requirements.

Trigger Sensitivity

The trigger sensitivities of the three circuits described were measured and compared in Fig. 17. The improvements in required amplitude and width of the trigger pulse are obvious from this comparison. The fact that d-c coupling was used in the circuits of improved trigger sensitivity does not mean to exclude the possibility of capacitive or inductive coupling. However, the trigger-sensitivity characteristic of the *2N68* flip-flop circuit indicates the need for a driver to supply a large, wide pulse. Although the widening of pulse width can be achieved by using monostable circuits, or the like, coupling this pulse to the low base input impedance is highly impractical.

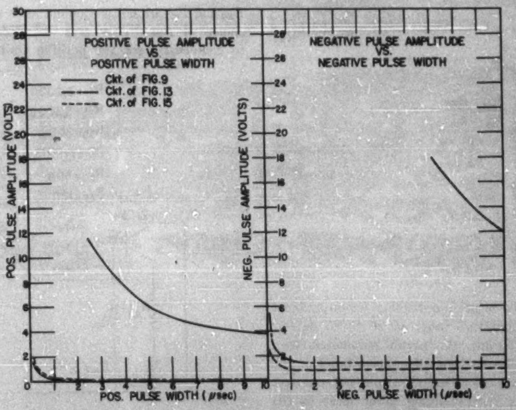


Fig. 17. Comparison of trigger sensitivities

Conclusion

The parameters of germanium transistors are known to be temperature-dependent. The effect of the temperature dependence of the transistor parameters on bistable power transistor circuits has been studied and a technique was developed to compensate for these effects. Based upon this technique, high-power bistable circuits have been designed to operate at elevated ambient temperatures with good d-c stability and trigger sensitivity. From the study it was found that the parameters most influencing stable operation of germanium transistors over wide temperature ranges are the d-c base resistance r_b, the saturation collector current I_{c0}, and the large signal amplification factor α.

Appendix I. Circuit Analysis

Upper Limit for R$_{1b}$ (Fig. 10)

When the transistor is in the OFF state, the emitter current and junction voltage are both near zero. By assuming the voltage across Ae to be zero, the following loop equations are obtained

$$I_{1b}\left(r_{1b}+\frac{R_{1b}R_{1f}}{R_{1b}+R_{1f}}\right)=\frac{-E_{bb}R_{1f}-E_{c,ON}R_{1b}}{R_{1b}+R_{1f}}$$

$$I_{1c}(r_{1c}+R_{1c})=E_{cc}+I_{c1}r_{1c}-\frac{E_{cc}R_{1c}}{R_{1c}+R_{2f}+r_2}$$

since

$$r_{1c}>>R_{1c},\ I_{c0}=I_{c1}+\frac{E_{cc}}{r_c}$$

and

$$\frac{E_{c,ON}}{r_{1c}}+I_{c0}>>\frac{E_{cc}R_{1c}}{(R_{1c}+R_{2f}+r_{2b})r_{1c}}$$

The condition $-I_b\geq I_c$ leads to equation 2.

Lower Limit for R$_{1b}$ (Fig. 11)

The loop equations are as given in the following:

$$E_{bb}=I_{f1}R_{1b}-I_{1b}(r_{1b}+R_{1b})$$

$$I_{1f}R_{f1}+(I_{2c0}+I_{1f})R_{2c}=E_{cc}I_{1b}r_{1b}$$

By assuming $I_{2c0}R_{2c}<<E_{cc}$ and solving for I_{1b},

$$I_{1b}=\frac{E_{cc}R_{1b}-E_{bb}(R_{1f}+R_{2c})}{r_{1b}R_{1b}+(r_{1b}+R_{1b})(R_{1f}+R_{2c})}$$

For $I_{c,ON}$ to be saturated

$$I_{1b}\geq\frac{E_c}{R_{1c}}\,\frac{1-\alpha}{\alpha}$$

Solving for R_{1b} and letting $\alpha/(1-\alpha)=\beta$, equation 4 is obtained.

Appendix II. Deviation of Upper and Lower Limits of the Base Resistances for a Pair of Flip-Flop Transistors with Two Cascaded Transistors (Fig. 13)

The limits for R_{1b} are the same as in Appendix I.

Upper Limits for R$_{2b}$ to Keep T-2 OFF

When *T-2* is OFF, the emitter current is very small; the base current is essentially the reverse collector current I_{2c0}. At the same time *T-3* is OFF and *T-1* is ON, the emitter voltage of *T-2* is $E_{3bb}-I_{3c0}R_{3b}$, while the collector voltage of *T-1* is $E_{1c,ON}$ (the saturation voltage of *T-1*).

By making use of the Thevenin's theorem, the base voltage of *T-2* can be shown to be

$$\frac{E_{2bb}R_{2f}}{R_{2b}+R_{2f}}-\frac{E_{1c,ON}R_{2b}}{R_{2b}+R_{2f}}-I_{2c,ON}\frac{R_{2b}R_{2f}}{R_{2b}+R_{2f}}$$

Thus, the voltage at the internal junction of *T-2* is

$$\frac{E_{2bb}E_{2f}}{R_{2b}+R_{2f}}-\frac{E_{1c,ON}R_{2b}}{R_{2b}+R_{2f}}-I_{2c,ON}\frac{R_{2b}R_{2f}}{R_{2b}+R_{2f}}-I_{2c0}r_{2b}$$

To keep *T-2* OFF, this voltage is required to be more positive than the emitter voltage of *T-2*, $(E_{3bb}-I_{3c0}R_{3b})$.

$$R_{2b}<\frac{R_{2f}(E_{2bb}-E_{3bb}+I_{3c0}R_{3b}-I_{2c0}r_{2b})}{E_{1c,ON}+E_{3bb}+I_{2c0}r_{2b}-I_{3c0}R_{3b}+I_{2c0}R_{2f}}$$

Lower Limit of R$_{2b}$ to Keep T-2 ON

When *T-2* is ON, *T-3* and *T-4* are both ON while *T-4* is OFF. The base voltage of *T-2* is, therefore $(I_{4b}r_{4b}+I_{3b}r_{3b}+I_{2b}r_{2b})$.

Thus, the current flowing through R_{2b} is

$$\frac{E_{2bb}+I_{4b}r_{4b}+I_{3b}r_{3b}+I_{2b}r_{2b}}{R_{2b}}$$

and the current flowing through R_{2f} is

$$\frac{E_{1cc}-I_{4b}r_{4b}-I_{3b}r_{3b}-I_{2b}r_{2b}}{R_{2f}+R_{1c}}$$

Consequently, the base current of *T-2* is

$$I_{2b}=\frac{E_{1cc}-I_{4b}r_{4b}-I_{3b}r_{3b}-I_{2b}r_{2b}}{R_{2f}+R_{1c}}-\frac{E_{2bb}+I_{4b}r_{4b}+I_{3b}r_{3b}+I_{2b}r_{2b}}{R_{2b}}$$

Solving for I_{2b}

$$I_{2b}=\frac{R_{2b}(E_{1cc}-I_{4b}r_{4b}-I_{3b}r_{3b})-(R_{2f}+R_{1c})(E_{2bb}+I_{4b}r_{4b}+I_{3b}r_{3b})}{[R_{2b}(R_{2f}+R_{1c})]+r_{2b}(R_{2f}+R_{1c}+R_{2b})}$$

Since the emitter voltage of *T-2* is $-(I_{4b}r_{4b}+I_{3b}r_{3b})$, the collector current of *T-2* is

$$I_{2c}=\frac{E_{2cc}-I_{4b}r_{4b}-I_{3b}r_{3b}}{R_{2c}}$$

To keep *T-2* ON, it is required that

$$I_{2b}>\frac{I_{2c}}{\beta_2}$$

Substituting and rearranging of term lead to the following equation:

$$R_{2b}>\left[\frac{E_{2bb}+I_{4b}r_{4b}+I_{3b}r_{3b}+(E_{2cc}-I_{4b}r_{4b}-I_{3b}r_{3b})\dfrac{r_{2b}}{R_{2c}\beta_2}}{\dfrac{E_{1cc}-I_{4b}r_{4b}-I_{3b}r_{3b}}{R_{2f}+R_{1c}}-\dfrac{E_{2cc}I_{4b}r_{4b}-I_{3b}r_{3b}}{R_{2c}\beta_2}\left(1+\dfrac{r_{2b}}{R_{2f}+R_{1c}}\right)}\right]$$

Upper Limit for R$_{3b}$ to Keep T-3 OFF

When *T-3* is OFF, *T-4* is also OFF. The internal junction voltage of *T-3*, $E_{3bb}-I_{3c0}(R_{3b}+r_{3b})$, is required to be more positive than the emitter voltage of *T-3*, $(E_{4bb}-I_{4c0}R_{4b})$.

Thus

$$R_{3b}<\frac{E_{3bb}-E_{4bb}}{I_{3c0}}+R_{4b}+r_{3b}$$

Lower Limit for R$_{3b}$ to Keep T-3 ON

Similar to the considerations described in the section entitled "Lower Limit of R_{2b} to Keep *T-2* ON," the base current of *T-3* can be derived to be

$$I_{3b}=\frac{R_{3b}(E_{2cc}-I_{4b}r_{4b})-R_{3c}(E_{3bb}+I_{4b}r_{4b})}{R_{3b}R_{3c}+r_{3b}(R_{3b}+R_{3c})}$$

and the collector current is

$$I_{3c}=\frac{E_{3cc}-I_{4b}r_{4b}}{R_{3c}}$$

Requiring $I_{3b}>I_{3c}/\beta_3$ and rearranging terms

$$R_{3b}>\left[\frac{E_{3bb}+I_{4b}r_{4b}+\left(\dfrac{E_{3cc}-I_{4b}r_{4b}}{R_{3c}\quad\beta_3}\right)r_{3b}}{\dfrac{E_{2cc}-I_{4b}}{R_{3c}}-\left(1+\dfrac{r_{3b}}{R_{3c}}\right)\dfrac{E_{3cc}-I_{4b}r_{4b}}{R_{3c}\beta_3}}\right]$$

Upper Limit for R$_{4b}$ to Keep T-4 OFF

When *T-4* is ON, its junction temperature is assumed to be ΔT higher than the ambient temperature. Immediately after *T-4* is turned OFF, the junction temperature will remain at $T+\Delta T$, where T is the ambient temperature. The internal junction voltage is $E_{4bb}-I_{4c0}(T+\Delta T)R_{4b}+r_{4b}(T+\Delta T)$ and the emitter voltage is zero. To keep *T-4* OFF, the internal junction voltage is required to be more positive than the emitter voltage, thus

$$R_{4b}<\frac{E_{4bb}}{I_{4c0}(T+\Delta T)}-r_{4b}(T+\Delta T)$$

Lower Limit for R$_{4b}$ to Keep T-4 ON

Similar to the considerations described in the section entitled "Upper Limits for R_{2b} to Keep *T-2* OFF," the base current of *T-4* can be shown to be

$$I_{4b}=\frac{R_{4b}E_{3cc}-R_{3c}E_{4bb}}{R_{4b}R_{3c}+r_{4b}(R_{3c}+R_{4b})}$$

and the collector current is as in the following:

$$I_{4c}=\frac{E_{4cc}}{R_{4c}}$$

Requiring $I_{4b}>I_{4c}/\beta_4$

$$R_{4b}>\left[\frac{\dfrac{E_{4cc}}{R_{4c}\beta_4}r_{4b}+E_{4bb}}{\dfrac{E_{3cc}}{R_{3c}}-\dfrac{E_{4cc}}{R_{4c}\beta_4}\left(1+\dfrac{r_{4b}}{R_{3c}}\right)}\right]$$

Appendix III. Derivation of the Upper and Lower Limits of the Base Resistance for a Pair of Flip-Flop Transistors with One Cascaded Transistor (Fig. 15)

This is a special case of Appendix II. As *T-4* is nonexistent, *I* etc. *R* etc. *r* etc. 0. Substituting these into the inequalities listed in equation 5 obtains those of equation 6. Note that *T-2* is a *2N68* power transistor in one case and a *2N34* in the other.

References

1. Some Aspects of the Design of Power Transistors, N. H. Fletcher. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, no. 5, May 1955.

2. Large Signal Behavior of Junction Transistors, J. J. Ebers, J. L. Moll. *Ibid.*, vol. 42, Dec. 1954, pp. 1761–72.

3. Temperature Compensated D.C. Transistor Amplifiers, E. Keonjian. *Ibid.*, Apr. 1954, pp. 661–71.

Analysis of Instability and Response of Reactors with Rectangular Hysteresis Loop Core Material in Series with Capacitors

JALAL T. SALIHI
STUDENT MEMBER AIEE

Synopsis: An exact analysis is given for a reactor having an idealized magnetization characteristic consisting of two linear regions of zero inductance, interrupted by a linear region of infinite inductance, when connected in series with a load comprising of capacitance and resistance under the action of a sinusoidal applied voltage. The recent developments in core materials have introduced magnetic materials such as Deltamax* and Orthonol† which have magnetization characteristic, at high values of saturation, approximating the ideal case. Thus the results obtained on the basis of an idealized analysis show good correlation with practical results obtained by experimenting on Deltamax cores. The circuit analyzed is the simplest circuit exhibiting an unstable and multivalued behavior attributed by several authors to a nonlinear resonance, sometimes referred to as ferroresonance phenomenon. Since the ideal magnetization characteristic assumed, consisting of regions of zero and infinite inductance, could not resonate with the finite load capacitance, the behavior of the circuit is explained by a different approach.

Analyzing the circuit on the basis of conduction angle is a new approach to the treatment of the problem. Thus a set of equations are obtained which are solved and expressed graphically. The co-ordinates of the curves are suitably chosen so that a single set of universal characteristics for different values of the phase angle of the load may be applied to any reactor, regardless of size or frequency, provided that the material of the core has magnetization characteristics similar to that of Deltamax. The multivalued behavior of the circuit and the degree of instability are determined by the charging rate of the capacitor which is controlled by the amount of resistance and capacitance present in the circuit. An explanation of the behavior of the circuit based on the effect of capacitor charge is given.

THE wide application of cores with rectangular hysteresis loop has increased the demand for an analysis of the behavior of such cores under the action of sinusoidal applied voltages. The behavior is similar in some respects to that of a gas tube. The similarity lies in the fact that the current for a given applied voltage flows only for a portion of the cycle, and is practically negligible outside this range. The conduction angle reduces with a reduction in the value of the voltage, reaching zero at a certain voltage called the critical voltage, beyond which the reactor ceases conducting. However, the relationship between conduction angle and applied voltage is quite different from that of a gas tube. It is controlled by two factors. the critical value of the flux density, at which the core saturates, and the nature of the load. In references 1 and 2 the behavior of such a reactor for cases of both pure resistance loads, and inductance-plus-resistance loads are investigated. For each case an analytical expression relating the conduction angle and applied voltage was deduced. The critical voltage at which the core starts firing is the same, regardless of the type of load. The firing angle characteristic is different for different loads. This difference becomes evident in the behavior of the circuit after firing. This may be seen in Fig. 1, where E_c represents the critical firing voltage and curves *a*, *b*, and *c* represent typical characteristics for the cases of resistive, inductive and capacitive loading respectively.

Nomenclature

A = effective core area (measured area × stacking factor of the core material) cm^2
B_c = critical value of flux density, gausses
C = load capacitance, farads
e = instantaneous value of applied voltage, volts
E = rms value of applied voltage, volts
E_c = critical value of rms applied voltage at which firing takes place, volts
E_c' = critical value of rms applied volts at which extinction takes place, volts
i = instantaneous value of current, amperes
I_{avg} = average value of current during half-cycle, amperes
N = number of turns of reactor winding
R_L = load resistance, ohms
R_w = resistance of reactor winding, ohms
R = effective resistance $R=R_L+R_w$, ohms
v = instantaneous value of voltage across capacitor, volts
V = voltage across capacitor during saturation, volts

Paper 56-10, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted June 20, 1955; made available for printing October 14, 1955.

Jalal T. Salihi is with the University of California, Berkeley, Calif.

The author gratefully acknowledges the contribution made to this study by Dr. A. M. Hopkin. Acknowledgment is also made to Dr. P. L. Chambré for his valuable suggestions.

---

* Arnold Engineering Company.

† Magnetics, Inc.

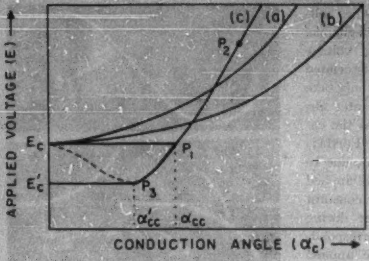


Fig. 1. Typical curves relating applied voltage and conduction angle for cases of resistance, inductive, and capacitive loading

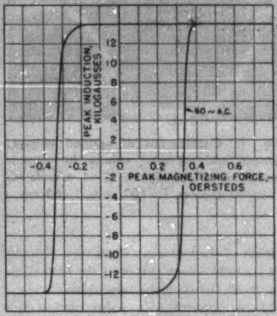


Fig. 2. Magnetization curve of Deltamax

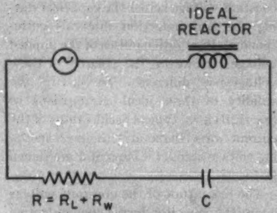


Fig. 3. Basic circuit analyzed with ideal reactor of zero resistance, having magnetization characteristic of Fig. 5

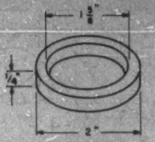


Fig. 4. Core of experimental Deltamax reactor, composed of 0.001-inch tape, and wound with 5,000 turns of no. 33 wire gauge having resistance of 145 ohms

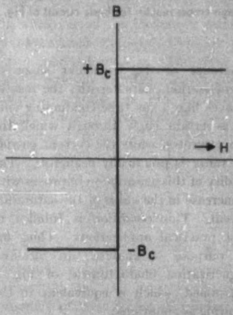


Fig. 5. Ideal magnetization characteristic assumed

q = instantaneous value of capacitor charge coulombs

Q = condenser charge during saturation, coulombs

α_f = angle for a positive half-cycle of the a-c supply voltage wave at which firing occurs, degrees or radians

α_c = angle for a positive half-cycle of the a-c supply voltage wave through which conduction is maintained, degrees or radians

α_{cc} = critical conduction angle corresponding to E_c, degrees or radians

α_{cc}' = critical conduction angle corresponding to E_c', degrees or radians

B = instantaneous value of flux density, gausses

ϕ = phase angle $= \tan^{-1}\dfrac{1}{\omega CR} = \tan^{-1}\dfrac{1}{\omega C(R_L+R_w)}$, degrees or radians

ϕ_c = critical value of phase angle where for $\phi<\phi_c$ the circuit does not exhibit unstable behavior, degrees or radians

Φ = total flux, lines

ω = angular velocity of supply voltage, radians per second

Curve *a* in Fig. 1 for the case of resistive loading is independent of the value of resistance, while curve *b* represents only one curve of a family corresponding to different values of $\phi=\tan^{-1}R/\omega L$ showing that the addition of inductance to the load will increase the conduction angle for the same applied voltage, thus making the reactor more sensitive to changes in voltage. Curve *c*, for the case of capacitive loading, represents only one of a family of curves corresponding to different values of $\phi=\tan^{-1} 1/\omega CR$ showing that for voltages ranging from $E_c'<E<E_c$ there is more than one possible mode of operation, depending on whether the voltage is increasing or decreasing. Thus starting from zero, if the voltage is increased the reactor does not conduct until the critical voltage E_c is reached, after which the operation point suddenly jumps to the point P_1. For further increase in voltage the conduction angle increases smoothly along the curve up to some point P_2. If the voltage is then reduced, the conduction angle reduces in the same manner down to the point P_1, where the voltage is again equal to E_c. A further reduction in voltage is described by the region P_1P_3 along which the conduction angle reduces smoothly to the point P_3, which corresponds to a second critical point at which a further reduction in voltage desaturates the core abruptly. This interesting nonlinear phenomenon associated with the case of capacitive loading is referred to in literature as the phenomenon of ferroresonance. It is important from theoretical and practical consideration and is made use of in control and computer circuits.

Investigations of the circuit treated in this paper but of different magnetization characteristic were made by several authors. The critical points were deduced either by graphical methods,[3] or from knowledge of the behavior of one reactor by applying the principle of similitude,[4] or by fitting certain equations to measured relations.[5] For the idealized magnetization characteristic assumed in this paper, exact mathematical relations are derived describing the behavior of the circuit for all ranges of operation including the critical points.

Theory and Assumptions

The experimental circuit consists of a nonlinear reactor with Deltamax core of magnetization characteristic shown in Fig. 2, connected to a 60-cycle-per-second (cps) variable voltage source in series, with a load consisting of the load resistance R_L and a capacitance C. The equivalent circuit is shown in Fig. 3, where the resistance of the reactor is lumped with the load. The design details of the reactor are given in Fig. 4.

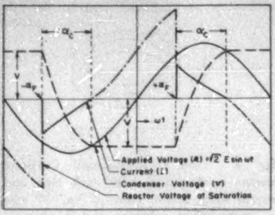


Fig. 6. Shapes and relative phases of applied voltage, current, capacitor voltage, and voltage across reactor for basic circuit of Fig. 3

Because of the particular shape of magnetization characteristic, the reactor behaves like a gas tube conducting only for a certain angle through which the core saturates, while the current outside this range is practically negligible. The validity of this assumption improves with an increase in the value of the saturation current. This condition is fulfilled in most practical applications. Thus, for the purpose of analysis, an idealized magnetization characteristic of Fig. 5 is assumed, which is equivalent to the following assumptions:

1. The current outside conduction range is zero.
2. The voltage across the capacitor during the nonconducting intervals remains constant. This follows from assumption 1, since the charge on a capacitor cannot change unless current flows.
3. The inductance of the core at saturation is zero.
4. Hysteresis and eddy current losses are negligible.

It is further assumed that the distributed and the winding capacitances of the reactor are negligible, as is the leakage of the load capacitor.

Analysis of Circuit

The wave shapes of current flowing in the circuit and the voltage across the capacitor will reveal some important relationships, shown in Fig. 6, where e represents the instantaneous value of the applied voltage. The origin of time is conveniently chosen so as to start from the initiation of a positive voltage wave. The instantaneous current, i flows only through the interval of time corresponding to the conduction angle α_c, and is zero outside the conduction interval because of the infinite inductance associated with the ideal reactor assumed. Thus, during nonconduction intervals, the instantaneous voltage v across the capacitor remains constant, provided the self-leakage of the capacitor is negligible. This is true since a change in voltage across the capacitor should be associated with a current flow which could only take place during the conduction intervals, during which the voltage across the capacitor reverses its polarity. Furthermore it may be concluded that, due to symmetry and periodicity of the applied voltage, the magnitude of the constant voltage V, across the capacitor during the nonconduction intervals corresponding to a positive pulse of the applied voltage, is equal to the magnitude of the constant voltage across the capacitor during the nonconduction intervals corresponding to negative pulses of the applied voltage. However, the polarity of these voltages is different. To justify the validity of these ideal assumptions to practical cases, typical oscillograms of the current wave shape and the corresponding voltage across the capacitor are shown in Fig. 7.

The magnitude of the constant voltage across the capacitor during nonconducting intervals may be calculated easily by making use of the fact that at $\omega t=\alpha_f+\alpha_c$, the current, i is zero. This may be seen in Fig. 6, which necessitates the condition that the voltage across the capacitor at this instant should be equal to the instantaneous value of the applied voltage, i.e.,

$$V=\sqrt{2}E\sin(\alpha_f+\alpha_c) \tag{1}$$

The behavior of the circuit may be represented by two differential equations describing the operation of the circuit during presaturation and saturation intervals. The differential equation corresponding to presaturation region is of the form

$$N\frac{d\Phi}{dt}\times 10^{-8}=V+\sqrt{2}E\sin\omega t \tag{2}$$

where the voltage drop across the resistance is assumed to be zero, and the constant voltage across the capacitor is substituted by an equivalent d-c source in series with the applied voltage.

During the saturation interval, the voltage drop across both resistance and capacitor must be accounted for while the voltage drop associated with the inductance of the reactor is zero. Thus the equation describing the operation of the circuit is of the form

$$Ri+\frac{1}{C}\int i\,dt=\sqrt{2}E\sin\omega t \tag{3}$$

These two equations give enough information to solve the circuit by a proper choice of the boundary conditions so as to fit the solutions together at the transition intervals between the two regions.

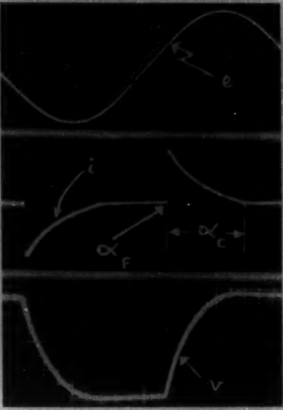


Fig. 7. Oscillograms of applied voltage, current, and capacitor voltage for experimental reactor of Fig. 4, C=1 microfarad, ϕ=70 degrees, E=100 volts, f=60 cps

It is important to point out that a relationship between applied voltage and average load current could not be expressed in a direct form. However, both these quantities could be expressed in terms of the conduction angle; thus an indirect relationship may be obtained. This is the reason for using the conduction angle as the basis for analysis.

Discussion of Relationships Derived

RELATIONSHIP BETWEEN FIRING ANGLE AND CONDUCTION ANGLE

The relationship between firing angle and conducting angle is derived in Appendix I, as equation 11, and is of the form

$$\alpha_f=\tan^{-1}\times\left(\frac{\sin\phi-\sec\phi\sin\alpha_c-\sin(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}{\sec\phi-\cos\phi+\sec\phi\cos\alpha_c+\cos(\alpha_c+\phi)\times\epsilon^{+\alpha_c\times\tan\phi}}\right) \tag{4}$$

showing that this relationship for a fixed $\phi=\tan^{-1}1/\omega CR$ is independent of the size of the reactor. That is, for two reactors of different sizes independently loaded, provided the phase angle in both cases is made equal by adjusting the loads and the conduction angles are made equal by adjusting the applied voltages, the firing angle will be the same for both cases.

RELATIONSHIP BETWEEN APPLIED VOLTAGE AND CONDUCTION ANGLE

The relationship between applied voltage and conduction angle is derived in Appendix II as equation 13 and is of the form

$$\frac{E}{AB_cN\omega}=\frac{\sqrt{2}}{10^8}\times\left(\frac{1}{(\pi-\alpha_c)\sin(\alpha_f+\alpha_c)-\cos\alpha_f-\cos(\alpha_f+\alpha_c)}\right) \quad (5)$$

where $E/AB_cN\omega$ is expressed as a function of both firing and conduction angles. This form of the relationship is more suitable for the purpose of calculation than the next expression, which is a function of α_c alone. This may be obtained by substituting equation 4 in equation 5 as shown in Appendix II, and the expression derived is of the form

$$\frac{E}{AB_cN\omega}=\frac{\sqrt{2}}{10^8}\left\{\frac{\sec\phi\times\tan^{-1}\left(\dfrac{\sin\phi-\sec\phi\sin\alpha_c-\sin(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}{\sec\phi-\cos\phi+\sec\phi\cos\alpha_c+\cos(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}\right)}{[(\pi-\alpha_c)\sin\alpha_c-\cos\alpha_c-1]+[(\pi-\alpha_c)\cos\alpha_c+\sin\alpha_c]\left(\dfrac{\sin\phi-\sec\phi\sin\alpha_c-\sin(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}{\sec\phi-\cos\phi+\sec\phi\cos\alpha_c+\cos(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}\right)}\right\} \quad (6)$$

From this it may be concluded that the relationship between $E/AB_cN\omega$ and the conduction angle α_c for a fixed value of $\phi=\tan^{-1} 1/\omega CR$, is independent of the size of the reactor; i.e., $E/AB_cN\omega$ should be equal for both of the two reactors of different sizes which are independently loaded, provided the conduction angle in both cases is made equal by adjusting the applied voltages.

RELATIONSHIP BETWEEN AVERAGE VALUE OF LOAD CURRENT AND CONDUCTION ANGLE

The relationship between average value of load current and oonduction angle is derived in Appendix III and of the form

$$\frac{RI_{avg}}{AB_cN\omega}=\left(\frac{2\sqrt{2}\cot\phi}{\pi}\right)\left(\frac{E}{AB_cN\omega}\right)\sin(\alpha_f+\alpha_c)=\frac{4\cot\phi}{\pi\times10^8}\left(\frac{\sin(\alpha_f+\alpha_c)}{(\pi-\alpha_c)\sin(\alpha_f+\alpha_c)-\cos\alpha_f-\cos(\alpha_f+\alpha_c)}\right) \quad (7)$$

where $RI_{avg}/AB_cN\omega$ is expressed as a function of both firing and conduction angle. This form of the relationship is more suitable for the purpose of calculation than the next expression, which is a function of α_c alone, obtained by substituting equation 4 in equation 7. This is shown in Appendix III, and the expression derived is of the form

$$\frac{RI_{avg}}{AB_cN\omega}=\frac{4\cot\phi}{\pi\times10^8}\left\{\frac{\sin\alpha_c+\cos\alpha_c\left(\dfrac{\sin\phi-\sec\phi\sin\alpha_c-\sin(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}{\sec\phi-\cos\phi+\sec\phi\cos\alpha_c+\cos(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}\right)}{[(\pi-\alpha_c)\sin\alpha_c-\cos\alpha_c-1]+[(\pi-\alpha_c)\cos\alpha_c+\sin\alpha_c]\left[\dfrac{\sin\phi-\sec\phi\sin\alpha_c-\sin(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}{\sec\phi-\cos\phi+\sec\phi\cos\alpha_c+\cos(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}\right]}\right\} \quad (8)$$

From which it may be concluded that the relationship between $RI_{avg}/AB_cN\omega$ and the conduction angle α_c for a fixed value of $\phi=\tan^{-1}$ $(1/\omega cR)$ is also independent of the size of the reactor.

RELATIONSHIP BETWEEN APPLIED VOLTAGE AND LOAD CURRENT

A direct relationship between rms applied voltage E and the average load current I_{avg} could not be obtained. However for any value of ϕ the corresponding values of $E/AB_cN\omega$ and $RI_{avg}/AB_cN\omega$ may be calculated by using equations 6 and 8. Thus, an indirect relationship between $RI/AB_cN\omega$ and $E/AB_cN\omega$ may be tabulated and expressed graphically. Furthermore, since both of these quantities for a given phase angle are independent of the size of the reactor, it follows that the relationship between them is also independent of the size of the reactor.

The fact that all the relationships derived are expressed in nondimensional form suggests a practical set of characteristics for different assumed values of phase angle ϕ. Calculation of these characteristics can be tedious because of the long equations involved; however, considerable time is saved if the following steps are taken, assuming any fixed value of ϕ:

1. From equation 4 find values of firing angle α_f, corresponding to different values of conduction angle α_c.
2. If values of α_c and the corresponding values of α_f are substituted in equation 1, values of $V/\sqrt{2}E$, corresponding to different values of α_c may be calculated.
3. If values of α_c and the corresponding values of α_f are substituted in equation 5, values of $E/AB_cN\omega$, corresponding to different values of α_c may be calculated.
4. Similarly, with the use of equation 7, values of $RI_{avg}/AB_cN\omega$ corresponding to different values of α_c may be calculated.
5. If the results obtained from steps 3 and 4 are tabulated, values of $RI_{avg}/AB_cN\omega$ corresponding to different values of $E/AB_cN\omega$ may be obtained.

Fixing values of ϕ at 10, 20, 30, 40, 50, 60, 70, and 80 degrees, this procedure was followed for each case. The calculated characteristics are expressed graphically in Figs. 8, 9, 10, 11, and 12.

Fig. 8 represents the relationship between firing angle and conduction angle, showing that as the load is made more capacitive, then for the same conduction angle, the corresponding firing angle is reduced. The straight line $\phi=0$ corresponds to the case of pure resistive load where $\alpha_f=180-\alpha_c$.

Fig. 9 shows the manner in which the constant voltage V across the capacitor during presaturation intervals expressed as a fraction of the maximum value $\sqrt{2}$ E of the sinusoidal applied voltage varies with conduction angle. The conduction angle varies as a result of the variation in applied voltage, as will be discussed next; thus, if the applied voltage is varied both V and α_c change while the relationship between $V/\sqrt{2}\ E$ and α_c is described by the curves of Fig. 9. A study of these curves is essential to explain the reason for instability; furthermore it will be shown that the degree of instability is related to the shape of these curves. For a fixed ϕ, as the conduction angle is increased due to an increase in applied voltage, V approaches nearer to the maximum value of the applied voltage; for example when $\phi=80$ degrees, $V/\sqrt{2}\ E$ is 0.76 and 0.98, corresponding to conduction angles of 30 and 60 degrees respectively. It may also be concluded that for a constant conduction angle, $V/\sqrt{2}\ E$ is larger for larger values of ϕ.

Figs. 10(A) and (B) show the variation of conduction angle with the nondimensional applied voltage $E/AB_cN\omega$, where $1/AB_cN\omega$ at a fixed frequency is a constant for a given reactor. When the voltage is gradually increased from zero, the reactor does not fire until a critical voltage E_c is reached. E_c is the value of E which satisfies the condition $E/AB_cN\omega=0.77\times10^{-8}$.

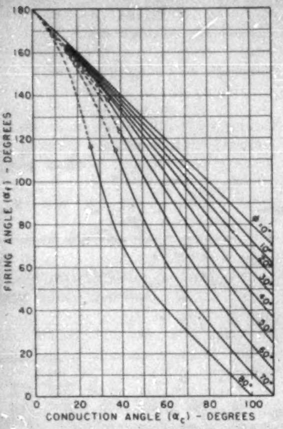


Fig. 8 (left). Calculated relationship between firing angle and conduction angle

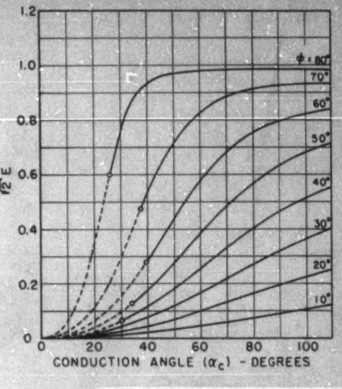


Fig. 9 (right). Calculated relationship between conduction angle and constant voltage across capacitor during presaturation intervals, expressed as a fraction of the maximum applied voltage

Figs. 11(A) and (B) show the variation of the nondimensional average current $RI_{avg}/AB_cN\omega$ with conduction angle, where $R=R_L+R_w$ is the effective resistance of the circuit.

Figs. 12(A) and (B) relate the average load current to applied voltage in a nondimensional form very convenient for practical application. In the range shown in Fig. 12(B) the relationship is remarkably linear.

In all the characteristics discussed the dashed lines correspond to negative resistance associated with unstable operation. The negative resistance is absent in those curves corresponding to $\phi<\phi_c$, where ϕ_c is a critical value for the phase angle.

Practical Application of Relationships Derived

The nondimensional set of characteristics given are determined by values of phase angle $\phi=\tan^{-1} 1/\omega CR$ assumed, while the effect of core size, winding details, the value of critical flux density, the frequency of supply, the applied voltage, the effective resistance, and the current flowing may be grouped and thought of as single factors $E/AB_cN\omega$ or $RI_{avg}/AB_cN\omega$, which do not change the shape but simply correspond to certain operating points on the characteristics; consequently they are applicable to any reactor size having Deltamax cores or any future cores which might be developed of similar character where the presaturation current is negligible. The characteristics are applicable at any such reasonably low frequency that the basic assumptions made are not violated.

To apply the characteristics to practical cases it is first necessary to know the value of $1/AB_cN\omega$, which at a fixed frequency is constant for a given reactor. This constant may be found either by calculation if design details are provided, or by measurements. Figs. 10(A) and 12(A) indicate that the critical firing voltage E_c is theoretically independent of the load and is given by that value of E at which $E/AB_cN\omega=0.77\times10^{-6}$. Thus $1/AB_cN\omega=0.707\times10^{-6}/E_c$ may easily be measured by a voltmeter connected across the applied voltage which is increased slowly up to the critical value at which firing takes place. The firing action may be noticed from observation of current oscillograms or by noting the sudden current rise in an ammeter connected in series with the load. Tests performed on the experimental reactor of Fig. 4 show that the values of $\phi<\phi_c$, where the circuit does not exhibit unstable modes, E_c is practically independent of the load condition as predicted theoretically, while for $\phi>\phi_c$, there is a slight reduction in the value of E_c, the deviation increases as ϕ is increased reaching 3 per cent for $\phi=80$ degrees. Thus for accurate measurement of E_c the phase angle should be less than ϕ_c or simply a pure resistive load may be chosen.

As a practical example, consider the reactor of Fig. 4 connected to a 60-cps 100-volt source, in series with a capacitor of 1 microfarad and a resistance of 820 ohms, which makes the effective resistance of the circuit $R=R_L+R_W=820+145=965$ ohms. It is required to find the average current, the conduction angle, and the firing angle.

The first step is to find the constants for the given reactor; thus a variable voltage was applied to the circuit while the capacitor was short-circuited so as to make the load resistive. The critical firing voltage measured was 43 volts; making use of this the constants are

$$\frac{1}{AB_cN\omega}=\frac{0.707\times10^{-6}}{43}=0.0164\times10^{-6}$$

$$\frac{R}{AB_cN\omega}=965\times0.0164\times10^{-6}=15.8\times10^{-6}$$

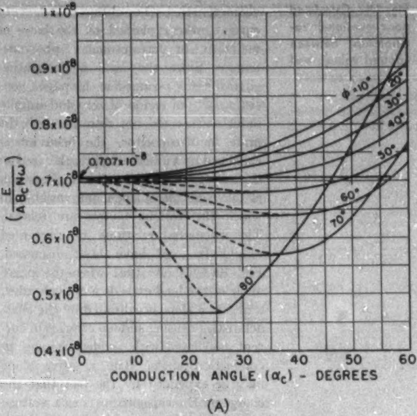
$$\phi=\tan^{-1}\frac{1}{\omega CR}=\tan^{-1}\frac{10^6}{2\pi\times60\times965}=70°$$

Now, for an applied voltage of 100 volts $E/AB_cN\omega=100\times0.0164\times10^{-6}=1.64\times10^{-6}$. The corresponding nondimensional average current is 0.50×10^{-6}, obtained from the characteristics of Fig. 12(B); thus $I_{avg}=0.50/10^6\times AB_cN\omega/R=0.50/15.8=0.0316$ ampere.

To obtain the conduction angle, use Fig. 10(B), from which $\alpha_c=91.5$ degrees. With this value of conduction angle the firing angle may be obtained from Fig. 8 from which $\alpha_f=21$ degrees.

The corresponding results obtained by direct measurements on the oscillograms of Fig. 7, are about 0.033 ampere, 95 degrees, and 23 degrees respectively.

To check the validity of the assumptions made in this paper a wider range of measurements was made on the experimental reactor. Fig. 13 shows a comparison of the calculated values of the average load current with experimental results, showing good correlation. Five widely different values of phase angle were chosen; in each case the applied voltage was varied and the corresponding average load current was obtained from measurement of oscillogram areas. The slight difference between measured and calculated results, the former being higher, is due to the flow of presaturation current



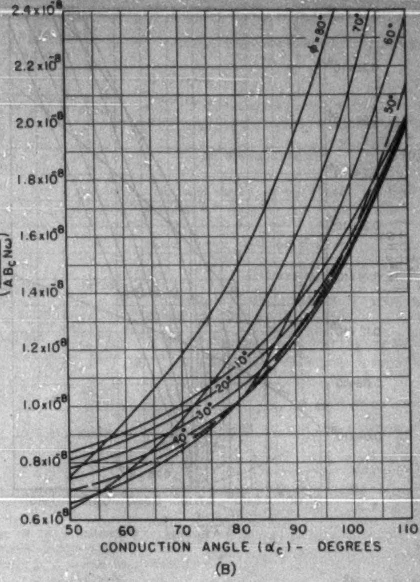
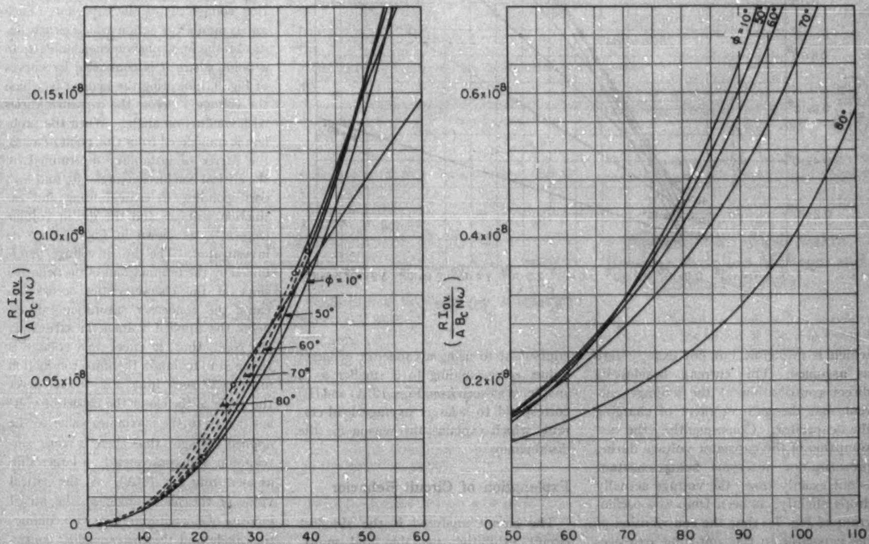


Fig. 10 (above). Calculated nondimensional relationship between applied voltage and conduction angle

Fig. 11 (below and right). Calculated nondimensional relationship between average current and conduction angle



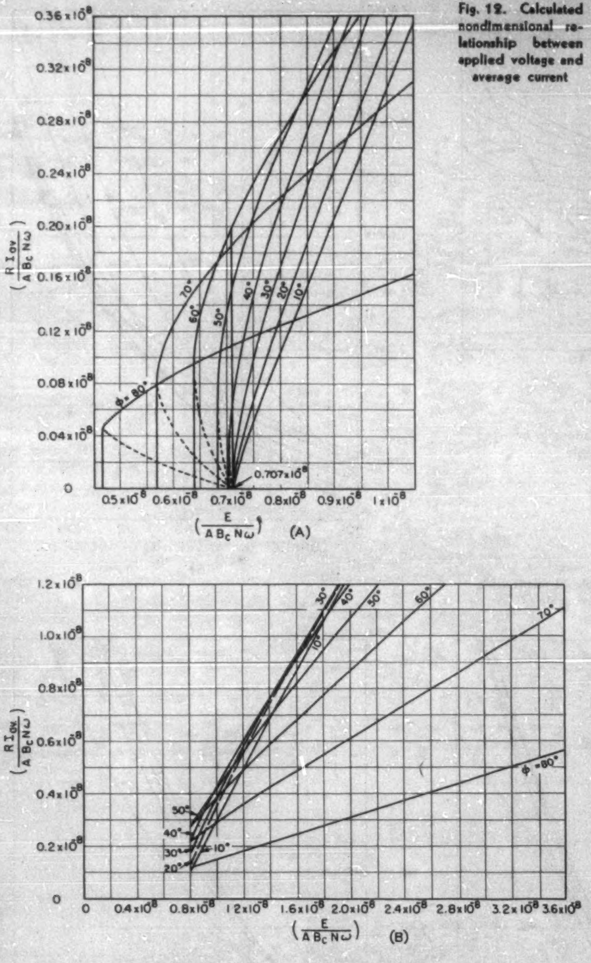


Fig. 12. Calculated nondimensional relationship between applied voltage and average current

which is negligible but not exactly zero as assumed. This current, besides its direct contribution to the average current, has the indirect effect of charging the capacitor. Consequently, the assumption of the capacitor voltage during presaturation intervals being constant is not exactly true; the voltage actually drops slightly, as seen from the oscillogram of Fig. 7; thus the contribution of the capacitor to the volt-time area necessary to raise the flux is slightly less than the theoretical case. This is roughly equivalent to using a capacitor of larger value, corresponding to a smaller ϕ; a smaller ϕ as seen from Figs. 12(A) and (B) correspond to a larger average load current, which explains the reason for the discrepancy.

Explanation of Circuit Behavior

The circuit analyzed is the simplest circuit exhibiting unstable and multivalued behavior. In analyzing the same circuit but with different magnetization characteristics, the abrupt changes of current were explained on the basis of nonlinear or ferroresonance phenomenon.[5,6] Since the ideal magnetization characteristic assumed in this paper, consisting only of region of zero and infinite inductance, could not resonate with the finite load capacitor, the behavior is explained by a different approach based on the variation of conduction angle, which is taken as an independent variable to which the various factors are related. The calculated relations are presented graphically and have been discussed. Fig. 10(A) shows that when the phase angle of the load exceeds a critical value, a negative slope is exhibited by the characteristics causing sudden changes in current associated with sudden changes in conduction angle. The negative slope may be explained by the fact that the charge on the capacitor acts as a voltage source helping the applied voltage to saturate the core. Thus at the instant when the critical firing voltage is just exceeded, the reactor fires and the capacitor is charged. The constant voltage V across the capacitor during presaturation intervals, Fig. 6, in helping the applied voltage will reduce the presaturation intervals; consequently, the conduction intervals are increased. The increase in conduction angle causes more current to flow which in turn raises the presaturation voltage across the capacitor. Such an accumulative action which clearly indicates the jump phenomenon builds up to a point where it is controlled by curves of Fig. 9, indicating the manner in which the voltage V across the capacitor varies with conduction angle. When the problem is considered from this point of view, the degree of instability determined by the critical conduction angles α_{cc} and α_{cc}', corresponding to critical firing and extinction voltages and the dip in voltage $(E_c - E_c')$, as shown in Fig. 1, may be investigated. The dip in voltage determined by the terminations of the negative slope of the characteristics occurs because the condenser saturation voltage helps the applied voltage in saturating the core; thus, in effect, less voltage is required to maintain the same conduction angle. As seen from Fig. 9 the larger the angle ϕ, the closer the capacitor voltage comes to the maximum value of the applied voltage; thus there is more contribution and consequently a larger dip, as seen from Fig. 10(A). At the critical value of the phase angle ϕ_c, the initial increase of capacitor voltage with conduction angle and the corresponding contribution to the volt-time area necessary to raise the flux is comparatively small, so that it is just neutralized by the effect of increase in the value of applied voltage. Hence for values of $\phi<\phi_c$ the circuit is stable. Regarding the critical conduction angle α_{cc} corresponding to the firing voltage, Fig. 10(A), shows that for values of $\phi>\phi_c$ this angle increases with ϕ reaching a maximum; it then reduces as the phase angle is further increased. In explanation of this action it may be seen that the angle α_{cc} is dependent on two factors:

1. The negative slope of the characteristics down to the point of minimum dip and the following positive slope. Both these slopes increase with the value of ϕ because of the larger capacitor voltage. Hence, assuming a fixed dip for the sake of argument, it may be seen that the larger these slopes, the smaller the angle α_{cc}.
2. If fixed slopes for the characteristics are assumed it may be seen that the larger the dip corresponding to a larger ϕ, the larger the angle α_{cc}.

If these two factors are considered simultaneously it may be concluded that as the value of ϕ is gradually increased beyond its critical value, the second effect predominates at first; thus the angle α_{cc} increases to a maximum point after which the first effect predominates; consequently α_{cc} reduces for further increase in ϕ.

If it is desired to find a relationship between α_{cc} and ϕ, substitute 0.707×10^{-8} for $E/AB_sN\omega$ in equation 6, while a relationship between the critical extinction angle α_{cc}' and ϕ may be obtained by differentiating equation 6 with respect to α_c and equating the resulting equation to zero.

Conclusions

The set of universal characteristics obtained on the basis of idealized assumptions show good correlation with practical measurement, obtained by experimenting on Deltamax reactors. Thus the behavior of such reactors and future reactors of similar magnetization character in series with capacitance may be predicted from the given characteristics, which are easy to use for design purposes, and very convenient since they may be applied to reactors of any size and at any such reasonably low frequency that the assumptions made are not violated. The analysis may be extended further to practical magnetic-amplifier circuits with capacitive loads, which will be treated in another paper. The analysis given, based on the variation of conduction angle, is a new approach to the problem of instability, giving a better physical picture to the peculiarity in the behavior of the circuit.

Appendix I. Relationship Between Firing Angle and Conduction Angle

The differential equation describing the operation of the circuit of Fig. 3 during saturation is of the form

$$Ri+\frac{1}{C}\int idt=\sqrt{2}\,E\sin\omega t$$

which may be written

$$\frac{di}{dt}+\frac{i}{CR}=\sqrt{2}\frac{E\omega}{R}\cos\omega t$$

Solution of this equation for the current gives

$$i=\frac{\sqrt{2}E}{\left(R^2+\frac{1}{\omega^2C^2}\right)^{1/2}}\sin\left(\omega t+\tan^{-1}\frac{1}{\omega CR}\right)+K\epsilon^{-\frac{1}{\omega CR}\times\omega t}$$

Letting $\phi=\tan^{-1}\frac{1}{\omega CR}$ in this equation gives

$$i=\frac{\sqrt{2}E}{R\sec\phi}\sin(\omega t+\phi)+K\epsilon^{-\tan\phi\times\omega t} \qquad (9)$$

The capacitor voltage at both instants when $\omega t=\alpha_f$ and $\omega t=\alpha_f+\alpha_c$ are equal. The polarity of this voltage is at $\omega t=\alpha_f$ to help the instantaneous applied voltage, while the reverse is true for the instant when $\omega t=\alpha_f+\alpha_c$; see Fig. 6. Thus at $\omega t=\alpha_f$

$$i=i\alpha_f=\frac{\sqrt{2}E}{R}\sin\alpha_f+\frac{V}{R}$$

Substituting this condition in equation 9 gives

$$\frac{\sqrt{2}E}{R}\sin\alpha_f+\frac{V}{R}=\frac{\sqrt{2}E}{R\sec\phi}\sin(\alpha_f+\phi)+K\epsilon^{-\tan\phi\times\alpha_f}$$

from which K may be evaluated and is given as

$$K=\left(\frac{\sqrt{2}E}{R}\sin\alpha_f+\frac{V}{R}-\frac{\sqrt{2}E}{R\sec\phi}\times\sin(\alpha_f+\phi)\right)\epsilon^{+\tan\phi\times\alpha_f}$$

Substituting this value of K back in equation 9 gives

$$i=\frac{\sqrt{2}E}{R\sec\phi}\left\{\sin(\omega t+\phi)+\left[\sec\phi\sin\alpha_f+\frac{V\sec\phi}{\sqrt{2}E}-\sin(\alpha_f+\phi)\right]\epsilon^{-\tan\phi(\omega t-\alpha_f)}\right\}$$

Substituting the second boundary condition $i=0$ at $\omega t=\alpha_f+\alpha_c$, shown in Fig. 6 in this equation, gives

$$0=\sin(\alpha_f+\alpha_c+\phi)+\left(\sec\phi\sin\alpha_f+\frac{V}{\sqrt{2}E}\sec\phi-\sin(\alpha_f+\phi)\right)\epsilon^{-\tan\phi(\alpha_f+\alpha_c-\alpha_f)}$$

from which

$$\epsilon^{+\alpha_c\times\tan\phi}\times\sin(\alpha_f+\alpha_c+\phi)=\sin(\alpha_f+\phi)-\sec\phi\sin\alpha_f-\frac{V}{\sqrt{2}E}\sec\phi$$

to expand

$$\epsilon^{+\alpha_c\times\tan\phi}[\sin\alpha_f\cos(\alpha_c+\phi)+\cos\alpha_f\sin(\alpha_c+\phi)]=\sin\alpha_f\cos\phi+\cos\alpha_f\sin\phi-\sec\phi\sin\alpha_f-\frac{V\sec\phi}{\sqrt{2}E} \qquad (10)$$

As was shown, $V=\sqrt{2}E\sin(\alpha_f+\alpha_c)$; see Fig. 6

therefore $\frac{V}{\sqrt{2}E}=\sin(\alpha_f+\alpha_c)=\sin\alpha_f\cos\alpha_c+\cos\alpha_f\sin\alpha_c$

Substituting this value of $V/\sqrt{2}E$ in equation 10 gives

$$\epsilon^{+\alpha_c\times\tan\phi}[\sin\alpha_f\cos(\alpha_f+\phi)+\cos\alpha_f\sin(\alpha_c+\phi)]=\sin\alpha_f\cos\phi+\cos\alpha_f\sin\phi-\sec\phi\sin\alpha_f-\sec\phi\sin\alpha_f\cos\alpha_c-\sec\phi\cos\alpha_f\sin\alpha_c$$

Dividing both sides of this equation by $\cos\alpha_f$ gives

$$\epsilon^{+\alpha_c\times\tan\phi}[\tan\alpha_f\cos(\alpha_f+\phi)+\sin(\alpha_c+\phi)]=\tan\alpha_f\cos\phi+\sin\phi-\sec\phi\tan\alpha_f-\sec\phi\cos\alpha_c\tan\alpha_f-\sec\phi\sin\alpha_c$$

Solving for $\tan\alpha_f$ gives

$$\text{Tan}\,\alpha_f=\left(\frac{\sin\phi-\sec\phi\sin\alpha_c-\sin(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}{\sec\phi-\cos\phi+\sec\phi\cos\alpha_c+\cos(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}\right) \qquad (11)$$

from which

$$\alpha_f=\tan^{-1}\times\left(\frac{\sin\phi-\sec\phi\sin\alpha_c-\sin(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}{\sec\phi-\cos\phi+\sec\phi\cos\alpha_c+\cos(\alpha_c+\phi)\epsilon^{+\alpha_c\times\tan\phi}}\right)$$

This equation relates the firing angle to the conduction angle for any fixed value of ϕ.

Appendix II. Relationship Between Applied Voltage and Conduction Angle

The differential equation describing operation of the circuit of Fig. 3 in the presaturation region is of the form

$$N\frac{d\Phi}{dt}\times10^{-8}=V+\sqrt{2}E\sin\omega t$$

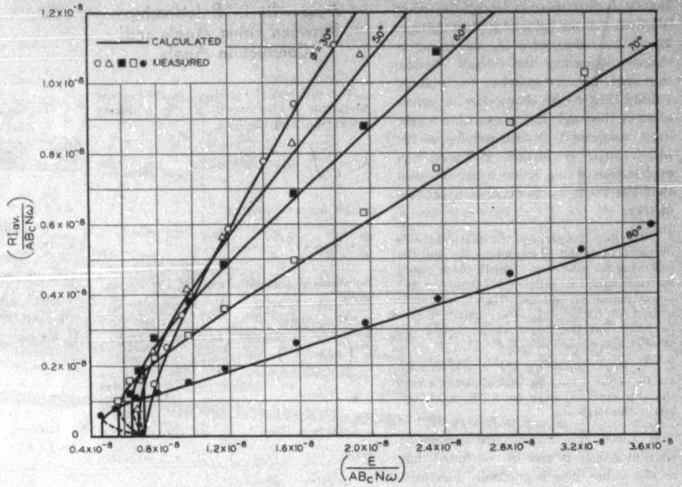
Since $d\Phi=Ad\beta$

therefore $d\beta=\frac{\sqrt{2}E\times10^8}{AN}\left(\frac{V}{\sqrt{2}E}+\sin\omega t\right)$

To integrate

$$\beta=K_1+\frac{\sqrt{2}E\times10^8}{AN}\int_{t_1}^{t}\left(\frac{V}{\sqrt{2}E}+\sin\omega t\right)dt \qquad (12)$$

Fig. 13. Comparison of measured and calculated relationship between applied voltage and average current expressed in nondimensional form. Measurements were made for experimental reactor of Fig. 4 connected in series with fixed capacitor of 1 microfarad and variable resistor adjusted to give appropriate values of phase angle, supplied from variable voltage source at 60 cps. Average load current estimated from areas of current oscillograms; measured $1/AB_cN\omega = 0.0164 \times 10^{-8}$



This equation simply says that the instantaneous value of flux density is equal to an initial flux density, plus a varying flux density proportional to the volt-time area of the applied voltage.

As seen from Fig. 6, the contribution of the flux to initiate a positive pulse of current beginning at $\omega t = \alpha_f$ starts the instant at which the preceding negative pulse of current drops to zero, i.e., at $\omega t = -[\pi - (\alpha_f + \alpha_c)] = \alpha_f + \alpha_c - \pi$. At this particular instant, the core is still saturated in the negative direction, thus $K_t = -B_c$. Substituting these two conditions in equation 12 gives

$$\beta = -B_c + \frac{\sqrt{2}E \times 10^8}{AN\omega}\int_{\alpha_f+\alpha_c-\pi}^{\omega t}\left(\frac{V}{\sqrt{2}E} + \sin \omega t\right) d\omega t = -B_c + \frac{\sqrt{2}E \times 10^8}{AN\omega} \times \left(\frac{V}{\sqrt{2}E}\,\omega t - \cos \omega t\right)_{\alpha_f+\alpha_c-\pi}^{\omega t}$$

At the instant when firing takes place $\beta = +B_c$ and $\omega t = \alpha_f$, thus

$$B_c = -B_c + \frac{\sqrt{2}E \times 10^8}{AN\omega} \times \left(\frac{V}{\sqrt{2}E}\,\omega t - \cos \omega t\right)_{\alpha_f+\alpha_c-\pi}^{\alpha_f}$$

Since $\cos(\alpha_f + \alpha_c - \pi) = -\cos(\alpha_f + \alpha_c)$

therefore $2B_c = \dfrac{\sqrt{2}E \times 10^8}{AN\omega} \times$

$$\left(\frac{V}{\sqrt{2}E}(\pi - \alpha_c) - \cos \alpha_f - \cos(\alpha_f + \alpha_c)\right)$$

from which

$$\frac{E}{AB_cN\omega} = \frac{\sqrt{2}}{10^8} \times \left(\frac{1}{\dfrac{V}{\sqrt{2}E}(\pi - \alpha_c) - \cos \alpha_f - \cos(\alpha_f + \alpha_c)}\right)$$

But $V = \sqrt{2}E \sin(\alpha_f + \alpha_c)$

$$\frac{E}{AB_cN\omega} = \frac{\sqrt{2}}{10^8} \times \left(\frac{1}{(\pi - \alpha_c)\sin(\alpha_f + \alpha_c) - \cos \alpha_f - \cos(\alpha_f + \alpha_c)}\right) \quad (13)$$

The relation between α_f and α_c is expressed by equation 11. When equation 11 is substituted in equation 13 the variable α_f may be eliminated and the voltage may be expressed in terms of a single variable α_c. This substitution may be performed directly; however, to obtain a compact solution, equation 13 may be written

$$\frac{E}{AB_cN\omega} = \frac{\sqrt{2}}{10^8}\left\{\frac{1}{[(\pi - \alpha_c)\sin \alpha_c - \cos \alpha_c - 1]\cos \alpha_f + [(\pi - \alpha_c)\cos \alpha_c + \sin \alpha_c]\sin \alpha_f}\right\}$$

Division of numerator and denominator by $\cos \alpha_f$ gives

$$\frac{E}{AB_cN\omega} = \frac{\sqrt{2}}{10^8}\left\{\frac{\sec \alpha_f}{[(\pi - \alpha_c)\sin \alpha_c - \cos \alpha_c - 1] + [(\pi - \alpha_c)\cos \alpha_c + \sin \alpha_c]\tan \alpha_f}\right\}$$

Substitution of equation 11 in this equation gives

$$\frac{E}{AB_cN\omega} = \frac{\sqrt{2}}{10^8}\left\{\frac{\sec \phi \times \tan^{-1}\left(\dfrac{\sin \phi - \sec \phi \sin \alpha_c - \sin(\alpha_c + \phi)\epsilon^{+\alpha_c \times \tan \phi}}{\sec \phi - \cos \phi + \sec \phi \cos \alpha_c + \cos(\alpha_c + \phi)\epsilon^{+\alpha_c \times \tan \phi}}\right)}{[(\pi - \alpha_c)\sin \alpha_c - \cos \alpha_c - 1] + [(\pi - \alpha_c)\cos \alpha_c + \sin \alpha_c]\left(\dfrac{\sin \phi - \sec \phi \sin \alpha_c - \sin(\alpha_c + \phi)\epsilon^{+\alpha_c \times \tan \phi}}{\sec \phi - \cos \phi + \sec \phi \cos \alpha_c + \cos(\alpha_c + \phi)\epsilon^{+\alpha_c \times \tan \phi}}\right)}\right\}$$

For any fixed value of ϕ this equation relates the applied voltage to the conduction angle where for a given frequency $\dfrac{1}{AB_cN\omega}$ is a constant of the reactor.

Appendix III. Relationship Between Average Value of Load Current and Conduction Angle

The magnitude of the voltage across the capacitor at both instants when $\omega t = \alpha_f$ and $\omega t = \alpha_f + \alpha_c$ are equal in magnitude, but opposite in polarity. This could be seen easily from Fig. 6. Consequently the same is true for the charge across the capacitor. The change in charge from $-Q$ at $\omega t = \alpha_f$ to $+Q$ at $\omega t = \alpha_f + \alpha_c$ is due to the current pulse flowing through the capacitor; this change may be expressed by

$$+Q = -Q + \frac{1}{\omega_c}\int_{\alpha_f}^{\alpha_f+\alpha_c} i\,d\omega t$$

Solving for Q gives

$$Q=\frac{1}{2\omega}\int_{\alpha_f}^{\alpha_f+\alpha_c} i\,d\omega t$$

Thus

$$V=\frac{Q}{C}=\frac{1}{2\omega C}\int_{\alpha_f}^{\alpha_f+\alpha_c} i\,d\omega t=\frac{\pi}{2\omega C}\times \frac{1}{\pi}\int_{\alpha}^{\alpha_f+\alpha_c} i\,d\omega t=\frac{\pi}{2\omega C}I_{\text{avg}}$$

from which

$$I_{\text{avg}}=\frac{2\omega C}{\pi}V=\frac{2\omega CR}{R\pi}V=\frac{2\cot\phi}{\pi R}V \tag{14}$$

But

$$V=\sqrt{2}E\sin(\alpha_f+\alpha_c) \tag{15}$$

Substituting equation 14 in equation 15 gives

$$I_{\text{avg}}=\frac{2\sqrt{2}}{\pi R}\cot\phi\times E\sin(\alpha_f+\alpha_c)$$

This equation may be put into a form more convenient for practical use, as

$$\frac{RI_{\text{avg}}}{AB_cN\omega}=\frac{2\sqrt{2}\cot\phi}{\pi}\,\frac{E}{AB_cN\omega}\sin[\alpha_f+\alpha_c)$$

Substitution of equation 13 in this equation gives

$$\frac{RI_{\text{avg}}}{AB_cN\omega}=\frac{4\times10^{-8}}{\pi}\cot\phi\times \left[\frac{\sin(\alpha_f+\alpha_c)}{(\pi-\alpha_c)\sin(\alpha_f+\alpha_c)-\cos\alpha_f-\cos(\alpha_f+\alpha_c)}\right] \tag{16}$$

The relation between α_f and α_c is expressed by equation 11. By substituting this relation in equation 16, α_f may be eliminated and the average current expressed in terms of a single variable α_c. The substitution may be performed directly; however, to obtain a more compact solution the following manipulation is convenient. Equation 16 may be written

$$\frac{RI_{\text{avg}}}{AB_cN\omega}=\frac{4\times10^{-8}}{\pi}\cot\phi\times \left[\frac{(\sin\alpha_f\cos\alpha_c+\cos\alpha_f\sin\alpha_c)}{(\pi-\alpha_c)(\sin\alpha_f\cos\alpha_c-\cos\alpha_f\sin\alpha_c)-(\cos\alpha_f\cos\alpha_c-\sin\alpha_f\sin\alpha_c)-\cos\alpha_f}\right]$$

Dividing numerator and denominator by $\cos\alpha_f$ gives

$$\frac{RI_{\text{avg}}}{AB_cN\omega}=\frac{4\times10^{-8}}{\pi}\cot\phi\times \left\{\frac{(\sin\alpha_c+\cos\alpha_c\times\tan\alpha_f)}{[(\pi-\alpha_c)\sin\alpha_c-\cos\alpha_c-1]+[(\pi-\alpha_c)\cos\alpha_c+\sin\alpha_c]\tan\alpha_f}\right\}$$

Substitution of equation 11 in this equation gives the equation shown in the following.

For any fixed value of $\phi=\tan^{-1}\dfrac{1}{\omega CR}$ this equation relates the average value of the load current to the applied voltage, where for a given frequency $1/AB_cN\omega$ is a constant of the reactor.

References

1. Analysis and Design of a Magnetic Frequency Multiplier, Otto J. M. Smith, Jalal T. Salihi. *AIEE Transactions*, vol. 74, Mar. 1955, pp. 99–106.
2. A Magnetic Frequency Multiplier, J. T. Salihi. *M.S. Thesis*, University of California, Berkeley, Calif., July 1954.
3. Critical Conditions in Ferroresonance, P. H. Odessey, Ernst Weber. *AIEE Transactions*, vol. 57, Aug. 1938, pp. 444–52.
4. Similitude of Critical Conditions in Ferroresonant Circuits, William T. Thomson. *Ibid.*, vol. 58, Mar. 1939, pp. 127–30.
5. Resonant Nonlinear Control Circuits, William T. Thomson. *Ibid.*, vol. 57, 1938, pp. 469–76.
6. Studies in Non-Linear Circuits, Chauncey Grey Suits. *Ibid.*, vol. 50, June 1931, pp. 724–36.
7. Self-Saturation in Magnetic Amplifiers, W. J. Dornhoefer. *Ibid.*, vol. 68, pt. II, 1949, pp. 835–50.
8. Magnetic Amplifiers (book), George M. Ettinger. John Wiley & Sons, Inc., New York, N. Y., 88 pp.
9. Mathematical Analysis of Nonlinear Circuits, A. Boyajian. *General Electric Review*, Schenectady, N. Y., vol. 34, Sept. 1931, pp. 531–37; Dec., pp. 745–51.

$$\frac{RI_{\text{avg}}}{AB_cN\omega}=\frac{4\times10^{-8}}{\pi}\cot\phi\left[\frac{\sin\alpha_c+\cos\alpha_c\left[\dfrac{\sin\phi-\sec\phi\sin\alpha_c-\sin(\alpha_c+\phi)e^{\alpha_c\times\tan\phi}}{\sec\phi-\cos\phi+\sec\phi\cos\alpha_c+\cos(\alpha_c+\phi)e^{\alpha_c\times\tan\phi}}\right]}{[(\pi-\alpha_c)\sin\alpha_c-\cos\alpha_c-1]+[(\pi-\alpha_c)\cos\alpha_c+\sin\alpha_c]\left[\dfrac{\sin\phi-\sec\phi\sin\alpha_c-\sin(\alpha_c+\phi)e^{\alpha_c\times\tan\phi}}{\sec\phi-\cos\phi+\sec\phi\cos\alpha_c+\cos(\alpha_c+\phi)e^{\alpha_c\times\tan\phi}}\right]}\right]$$

◆

Discussion

George C. Feth (Carnegie Institute of Technology, Pittsburgh, Pa.): The author's core idealization introduces a convenient simplification in the problem, that is hereby reduced to one of solving numerically certain transcendental equations and presenting the results systematically. However, some clearer indication would be desirable concerning the range of the pertinent circuit parameters over which the assumptions, and thus also the results, can be applied.

The idealization used neglects both the width of the hysteresis loop and its slope in the saturated region. In the operation of a circuit using such a rectangular loop reactor and a series load containing a constant resistance, inductance, or capacitance the magnetizing current which flows while the core is unsaturated causes voltage drops in the load. For small values of capacitance or large values of resistance or inductance these drops may be appreciable in comparison with the saturation voltage of the core; thus the critical supply voltage at which saturation of the core first occurs is not independent of the load used.

Moreover, when capacitive loads are used, as the supply voltage is increased above the critical voltage, a kind of resonance or exchange of energy may occur between the capacitor and the nonlinear saturated inductance of the core. The "resonant frequency" of this mode of operation is dependent upon voltage as well as the circuit parameters; several unidirectional peaks of current may occur in a single half-cycle.

Moreover, the operation of the circuit may no longer be periodic on one cycle of supply voltage but may contain subharmonics or may not be periodic at all. These oscillations of current may even swing the core back into the unsaturated region.

These phenomena, which have been recognized for poorer grade, less "square," core materials, also occur with the newer materials. For instance, with an Orthonol core of 0.002-inch tape, inside diameter 1¹/₄ inches, outside diameter 1³/₄ inches, height 1 inch, all of these phenomena were observed on a 200-turn winding at 60 cycles with series capacitance in the range 0.1 to 10 microfarads. With much larger values of capacitance these effects are less noticeable, as would be expected.

Jalal T. Salihi: I wish to thank Mr. Feth for bringing up some points of interest along which further investigations may be made. It is well recognized that the width of the hysteresis loop and its slope imposes limitations on the validity of the idealized analysis presented in this paper. However, for reactors using core materials such as Deltamax and Orthonol there is a wide range of parameters over which the idealized assumptions are closely met. This is because such core materials have a narrow width of hysteresis loop; a high value of presaturation inductance; and a low value of saturation inductance, compared with less nearly perfect core materials. Consequently, for a given phase angle the resistance of the circuit may be increased appreciably without causing noticeable voltage drop in the load associated with the magnetization current. Also, the value of load capacitance may be reduced appreciably before its capacitive reactance reaches a value comparable with the inductance of the core to cause interaction and consequent oscillation. Thus, for reactors using such cores the limitation imposed on the analysis of operation by the maximum value of load resistance and the minimum value of capacitance leaves a

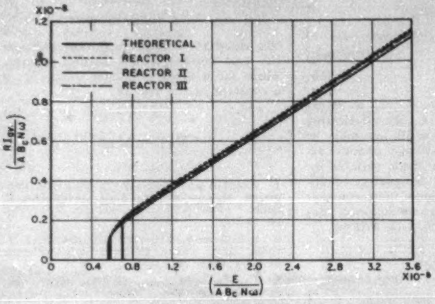


Fig. 14. Comparison of theoretical results with practical measurements made on three reactors of different sizes

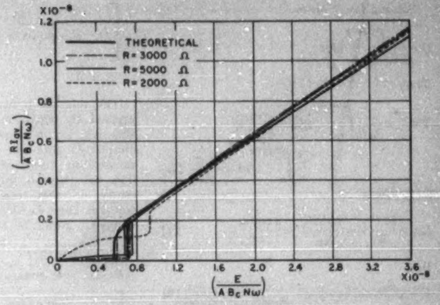


Fig. 15. Limitation imposed by magnetization current on validity of idealized analysis

Table I. Sizes and Core Materials of Reactors

Reactor No.	Material	Inside Diameter, Inches	Outside Diameter, Inches	Height, Inches	No. of Turns	Firing Voltage	$\frac{1}{AB_cN\omega}$
1	Orthonol *50026-1A*	1.0	1.5	0.375	1,200	22	0.0321×10⁻
2	Orthonol *50001-2A*	1.5	2.5	0.500	500	27	0.0262×10⁻
3	Deltamax *5320-D2*	1.5	2.5	0.500	600	32	0.0221×10⁻

Table II. Minimum Capacitance and Various Resistances of Experimental Circuit

R, Ohms	C, Microfarads	Phase Angle, Degrees
145	1.5	85°18′
200	1.1	85°30′
300	0.7	85°30′
500	0.35	86°12′
1,000	0.15	86°48′
2,000	0.075	86°48′
4,000	0.035	87°

wide range of operation satisfying the idealized assumptions. It may also be seen that during this range the core is well utilized as far as power output is concerned, while high values of resistance and low values of capacitance cause a reduction in power output.

A comparison of theoretical results with practical measurements made on three reactors of different sizes is given in Fig. 14, showing good agreement. The size and core material of the reactors used are as in Table I. The effective resistance $R=R_w+R_L$ in series with the reactors are 250, 100, and 100 ohms respectively, while the phase angles were fixed at 70 degrees. The values of $1/AB_cN\omega$ for each reactor was measured by the method described in this paper and are tabulated. The average value of load current was read from a d-c ammeter supplied by a bridge rectifier composed of silicon rectifiers (*1N 338*). The correlation between theory and experiment seen in Fig. 14 is a typical example and may be obtained for a wide range of parameters.

To study the limitation imposed on the value of parameters for good check with theory, the effect of both the width of the hysteresis loop and its slope on the behavior of the circuit are considered. The effect of each may be considered separately as follows.

Limitation Due to Width of Hysteresis Loop

The limitation imposed by the magnetization current on the validity of the idealized analysis may be seen from Fig. 15, where theoretical characteristic for $\phi=70$ degrees is compared with three measured characteristics obtained by measurements on the experimental reactor of Fig. 4. The measured characteristics correspond to three widely different values of resistances of 3,000, 5,000, and 20,000 ohms, and a fixed value of phase angle of 70 degrees. From examination of Fig. 15, it may be concluded that for the given reactor, and assuming a phase angle of 70 degrees, the effect of magnetization current on the behavior of the circuit is almost negligible up to a load resistance of 3,000 ohms. Thus for $\phi=70$ degrees and $R\leq 3{,}000$ ohms the firing and extinction voltages are practically independent of the load. At higher values of load resistance the effect of magnetization current appears, causing a change in the firing characteristic. Thus, as the load resistance is increased, the firing and the extinction voltages increase and the separation between them reduces till they become hardly distinguishable at a load resistance $R=20{,}000$ ohms. The jump in current may be eliminated completely at higher values of resistance.

Knowing the behavior of one reactor at a certain phase angle, we can predict the behavior of other reactors from the following nondimensional consideration. This holds true for the cases corresponding to $R\leqq 3{,}000$ for the experimental reactor, where, as seen from Fig. 15, the effect of the magnetization current is noticeable but not large enough to disturb the nondimensional relationship between $E/AB_cN\omega$ and $RI_{avg}/AB_cN\omega$. Thus, let two reactors have identical operation, i.e., having equal values of conduction angle and phase angle, and let I_m/I_{avg} denote the ratio of magnetization current to the average value of saturation current at the given conduction angle. Making this ratio the same for both reactors gives

$$\frac{I_{m1}}{I\text{avg}_1}=\frac{I_{m2}}{I\text{avg}_2}$$

This may be written as

$$\frac{\dfrac{R_1I_{m1}}{A_1N_1B_{c1}\omega_1}}{\dfrac{R_1I\text{avg}_1}{A_1N_1B_{c1}\omega_1}}=\frac{\dfrac{R_2I_{m2}}{A_2N_2B_{c2}\omega_2}}{\dfrac{R_2I\text{avg}_2}{A_2N_2B_{c2}\omega_2}}$$

Since the conduction angles of both reactors are equal, the denominator of both sides of this equation are equal, thus

$$\frac{R_1I_{m1}}{A_1N_1B_{c1}\omega_1}=\frac{R_2I_{m2}}{A_2N_2B_{c2}\omega_2} \tag{17}$$

But

$$I_m\alpha\frac{H_cl}{N} \tag{18}$$

Thus, from equations 17 and 18

$$\frac{R_2}{R_1}=\frac{A_2N_2^2B_{c2}\omega_2l_1H_{c1}}{A_1N_1^2B_{c1}\omega_1l_2H_{c2}} \tag{19}$$

This equation gives the condition of equal ratios of saturation to magnetization current at equal values of phase angles and equal conduction angles, for two reactors of different sizes. For two reactors using the same core material and the same supply frequency

$$R_2=\frac{A_2N_2^2l_1}{A_1N_1^2l_2}R_1$$

Applying this equation to reactor 3, and the experimental reactor of Fig. 4 having a resistance of 3,000 ohms in the load, gives

$$R_3=\frac{(1.37)(600)^2\times 16}{(0.24)(5{,}000)^2\times 14.4}3{,}000=274\text{ ohms}$$

Thus, it may be concluded that at a phase angle of $\phi=70$ degrees, the effect of magnetization current on the behavior of reactor 3 is almost negligible for values of $R<274$ ohms. At $R=274$ ohms the deviation from the idealized analysis corresponds to the curve $R=3{,}000$ of Fig. 15.

LIMITATION DUE TO SLOPE OF MAGNETIZATION CHARACTERISTIC

If the value of capacitance in the circuit is made sufficiently small, then as the applied voltage is increased the circuit undergoes a chain of irregular modes of operation as described by Mr. Feth. However, measurement on the experimental reactor of Fig. 4 indicates that such a phenomenon takes place only when the load is so highly capacitive that the phase angle exceeds 85 degrees. Thus, for all values of $0<\phi<85$, the idealized analysis controls the operation of the reactor considered. This conclusion was arrived at from an observation of the experimental circuit where the value of load resistance was fixed and the behavior of the circuit was observed as $E/AB_cN\omega$ was increased from zero to 3.6×10^{-8}, at different settings of the capacitor. For values of capacitance sufficiently small the behavior is characterized by the appearance of additional pulses of current in a single half-cycle, which may increase in number as the applied voltage is increased. Up to four of these pulses were noticed. In the range of the applied voltage extending between two of the pulses, oscillation and other phenomena may occur. For a given value of load resistance the number of these pulses is reduced as the value of load capacitance is increased. Consequently, for a given range of applied voltage there is a minimum value of load capacitance at which the appearance of the pulses may be eliminated. For the experimental circuit the minimum value of capacitance corresponding to different values of resistance in the circuit was measured; see Table II. Thus, the maximum value of $\phi=\tan^{-1} 1/\omega CR$ is practically constant, and independent of load.

From these results the following conclusions may be made regarding the operation of the experimental circuit:

1. There is only a narrow range of phase angle $85<\phi<90$ degrees where the idealized analysis in this paper fails, and where subharmonic oscillation and other nonlinear phenomena may occur.
2. The restriction on the minimum value of load capacitance is not imposed by a resonance phenomenon, but rather by the value of load resistance such that ϕ remains constant at about 85 degrees. The nonlinear phenomena described occur at smaller values of capacitance.
3. The statement made by Mr. Feth that the nonlinear phenomena described are due to a kind of resonance or exchange of energy between the capacitor and the nonlinear saturated inductance does not seem to be justifiable. Reactors using Deltamax and Orthonol cores are known for their small values of saturated inductance. For example, in reference 2, the saturation reactance of the experimental reactor of Fig. 4 is estimated at about 10 ohms at 60 cycles per second. This is hardly noticeable when compared with the values of capacitive reactance at which these phenomena occur. Consequently it is more reasonable to associate the phenomena with higher values of inductance, such as the presaturation inductance. Values of capacitance associated with capacitive reactances comparable to the saturation inductance are of the order of several hundreds of microfarads. No investigation was made at such high values of capacitance, where there might be the possibility of nonlinear oscillations. This may impose limitation on the maximum value of capacitance which may be used in the range where the idealized analysis is applicable. However, it may be seen that the range of phase angle at which the capacitance reaches such high values is only a few degrees even at the minimum value of resistance determined by the winding resistance of 145 ohms for the experimental reactor considered.

Low-Impedance Operational Characteristics of Toroidal Cores Used in Magnetic Amplifiers

C. L. BOYAJIAN
ASSOCIATE MEMBER AIEE

Synopsis: The need for knowing the low-impedance operating characteristics of a toroidal core or cores to be used in a magnetic amplifier operating under low-impedance control conditions is shown in this paper. A method is then presented for determining these low-impedance characteristics without first winding turns on the core. This method is presented for the edification of persons who are working on the development of new core-testing techniques.

Encouraging results of a preliminary investigation of this low-impedance testing method are shown.

MANY magnetic amplifiers approach low-impedance control operation. For this reason, it is desirable to have a method for testing cores used in these amplifiers which will simulate low-impedance control operation.

The testing method to be used must be so designed to predict satisfactorily the characteristics of the core for such operation without the need for first winding a large number of turns on the core. Most testers presently being used, which fulfill the latter requirement by using multiple-turn connectors or similar test jigs, operate on a high-impedance control principle. The most commonly used methods are the constant-current reset test method,[1,2] and the sine-current test method.

These methods produce good results for high-impedance-control magnetic-amplifier applications, but the results are not as good for use with low-impedance-control magnetic amplifiers. A more accurate determination of the operational characteristics of toroidal cores used in low-impedance control operation is desirable for applications requiring more accurate matching of cores than now possible.

Data comparing the core characteristics measured on a constant-current reset core tester with those measured on an actual low-impedance-control magnetic amplifier, confirmed the need for a method to determine the low-impedance operating characteristics of a core.

Determination of Mode of Operation of a Magnetic Amplifier

The need for a low-impedance-type test method was also made apparent by the results of an investigation to determine the type of impedance control operation of a magnetic amplifier.

The power gain and/or drift requirements of many low-power magnetic-amplifier applications require that the amplifier time constant be approximately 1/100 seconds or longer. This applies to applications operating at supply frequencies in the vicinity of 400 cycles per second (cps) or less.

The mode of operation of a full-wave

Paper 56-220, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 19, 1955; made available for printing December 15, 1955.

C. L. BOYAJIAN is with General Electric Company, Schenectady, N. Y.

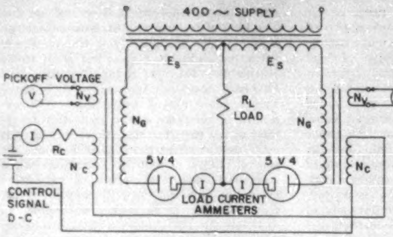


Fig. 1. Full-wave center-tapped circuit used for measuring amplifier control characteristics. Electronic tubes used for ideal rectifiers

center-tapped magnetic amplifier designed with a time constant of 1/100 seconds was investigated to determine the type of impedance control operation under which this amplifier can be classified.

Method of Obtaining Data

A full-wave center-tapped circuit as shown in Fig. 1 was used for the investigation. Two types of materials were investigated: Deltamax, (50-per-cent (%) nickel-iron grain-oriented 2-mil tape) and Supermalloy (79% nickel, 5% molybdenum, 16% iron, 2-mil tape). The size of the Supermalloy and Deltamax cores used is 1.0 inch inside diameter (ID) by 1.375 inches outside diameter (OD) by 0.25 inch. Two matched cores of each type, with average characteristics, were selected and wound for the full-wave center-tapped circuit. Electronic diodes were used instead of metallic rectifiers to eliminate the leakage current and capacitance effects. Only 400-cps operation was investigated.

To determine the effect of the control circuit impedance on the operation of the amplifier, the following conditions of control impedance were observed: (a) low impedance, an N^2/R ratio$=1{,}000$, with as low a control resistance as practically possible to allow the free flow of the induced harmonic currents (the N^2/R ratio is the ratio of the control turns squared to the total control circuit resistance. This ratio is used as a figure to represent the type of impedance control operation of the magnetic amplifier); (b) intermediate impedance, resulting in an amplifier time constant of 1/100 seconds—for the Deltamax cores, $N^2/R=160$, for the Supermalloy cores, $N^2/R=40$; (c) high impedance, with an N^2/Z ratio approaching zero obtained by using a 10-henry choke in the control circuit to oppose the flow of induced harmonic currents so as to maintain a steady d-c control current (the Z is the total impedance of the control circuit); (d) in addition, a control impedance condition with $N^2/R=10$ was investigated. This condition falls between the high- and intermediate-impedance operation but approaches the high-impedance control condition.

Results of Gain Curve Comparisons

Amplifier gain curves were taken for each of the afore-mentioned conditions. The results are shown in Figs. 2 and 3. The results of Fig. 2 show that with Deltamax, there is an appreciable shift to the right as the control impedance is decreased. The gain also decreases in this direction. The important observation, however, is the closeness of the intermediate- and low-impedance gain curves. This seems to indicate that the intermediate-impedance operation approaches the low-impedance operation.

Fig. 3 shows an entirely different characteristic for Supermalloy. As the control impedance is decreased, the gain curve in the linear region shifts to the right and then returns to the left. The intermediate impedance results in the gain curve furthest to the right. From this, one might conclude that the intermediate-impedance condition lies between the high and low cases. No definite conclusion can be made, however.

The magnitude of the supply voltage used to take these gain curves was slightly less than the value that would be equivalent to the voltage induced by a complete flux swing of the core from minus saturation to plus saturation.

The plots of Figs. 2 and 3 show a high cutoff load current. This load current could not be reduced any further because of the high magnetizing current which was necessary as a result of the relatively small number of turns on the gate winding.

Comparison of Wave Forms

The investigation was continued by observing the actual control current and reset voltage wave forms for the conditions of control circuit impedance mentioned previously by use of the same cores. The additional condition of control circuit impedance between the intermediate and high conditions with an N^2/R ratio of 10 was again observed for Supermalloy and Deltamax cores. This was done to show the transition between the intermediate- and high-impedance case since there was a considerable difference. These wave forms are shown in Figs. 4–7. Three firing angles are shown.

The reset voltage is the negative half-cycle in Figs. 5 and 7. The superimposed sine wave in Fig. 5 is for reference only. The sweep as shown is from left to right.

A highly grain oriented nickel-iron core material such as Deltamax exhibits square *B-H* loop properties when a sinusoidal voltage is applied to the core as in low-impedance operation. The reset current for Deltamax, under these low-impedance operating conditions, approaches a square shape as shown in Fig. 4(A). The theoretical explanation of this characteristic wave form is presented in reference 3. Supermalloy, which has a rounded *B-H* loop under low-impedance conditions, would cause a low-impedance reset current wave form that is rounded as shown in Fig. 6(A).

The reset voltage wave form is sinusoidal for the low-impedance operation of both materials as is seen in Figs. 5(A) and 7(A).

As the control circuit impedance is increased, the control current wave shape approaches a constant direct current as shown in Figs. 4(D) and 6(D). The reset voltage is no longer a sine wave but takes on a shape determined by the high-impedance control *B-H* characteristics of the core as seen in Figs. 5(D) and 7(D).

From a comparison of the wave shapes photographed under the intermediate-impedance conditions [Figs. 4(B), 5(B), 6(B), and 7(B)] with those taken under the low-[Figs. 4(A), 5(A), 6(A), 7(A)] and high-[4(D), 5(D), 6(D), 7(D)] impedance conditions, one can conclude that the intermediate case approaches the low-impedance case.

Conclusion

The comparisons of the gain curves and the current and voltage wave forms have shown that this low-power magnetic amplifier with a time constant of 1/100 seconds operates under a condition approaching low-impedance control operation. From these results, it may be concluded that many full-wave magnetic amplifiers operating at supply frequencies of 400 cps or less and designed to have a time constant of 1/100 seconds or less operate under conditions approaching low-impedance-control operation.

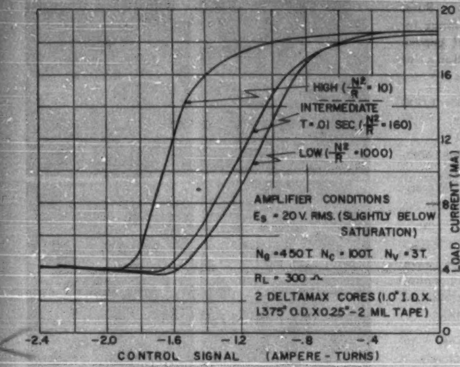


Fig. 2. Amplifier control characteristic curves showing the effects of varying the control circuit impedance on the current gain curves of Deltamax cores in a full-wave center-tapped circuit using electronic diodes

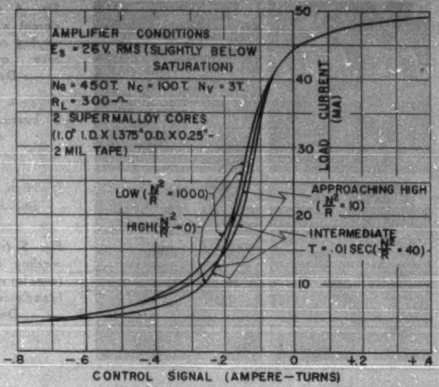


Fig. 3. Amplifier control characteristic curves showing the effects of varying the control circuit impedance on the gain curves of Supermalloy cores in a full-wave center-tap circuit using electronic diodes

In light of this information, the need for a method to determine the low-impedance operational characteristics of a toroidal core is realized.

Low-Impedance Core-Test Method

The work presented in this paper was done in conjunction with an investigation to determine the best methods for testing toroidal cores used in magnetic amplifiers. One of the commonly used core-testing methods, the constant-current reset tester,[1,2] was thoroughly investigated to determine the correlation existing between toroidal-core operating characteristics as measured on the core tester and in an actual full-wave center-tapped magnetic amplifier circuit. A comparison of core gain measurements of Deltamax cores (1-inch ID by 1.375-inch OD by 0.25-inch 2-mil tape) taken with the constant-current reset core tester with those taken with the magnetic amplifier under low-impedance control operation showed poor correlation; see Fig. 8. The method used to take these data will be explained later. Good correlation was not expected under these conditions, however, since the constant-current reset core tester operated under high-impedance control conditions. A comparison of the same core-tester gain data with the gain measurements taken with the magnetic amplifier under high-impedance control operation is shown in Fig. 9. This plot shows the correlation that is presently achieved using this constant-current reset core tester.

In addition, there is no guarantee that cores matched under high-impedance control conditions will be matched when operating under low-impedance control conditions. For many applications, this change in the match is tolerable. However, many applications require an accurate match or measurement of the core characteristics for low-impedance operation.

PROBLEM

The low-impedance core tester must simulate the low-impedance control operation of the amplifier and satisfactorily predict the characteristic of the core for such operation without first winding a large number of turns on the core. To achieve low-impedance operation, a sinusoidal voltage must be applied to the core. A sinusoidal voltage is possible only if the total impedance of the supply circuit, as seen from the turns of the test jig, is very low.

A toroidal core tester, if it is to be of practical value when it tests a large number of cores, must use a multiple-turn connector or test jig of some type to make the required turns around the core for the excitation and pickoff signals. The test jig would eliminate the necessity of winding turns on the core. The physical size of such a test jig greatly limits the number of allowable turns for small-size cores.

A major problem is to maintain the resistance of the test circuit low enough with the few test turns of the connector to achieve a sine-voltage excitation. Most connectors develop increasing contact resistance with use if not continuously cared for. For example, the impedance of the circuit presented in this paper must be of a value to result in a ratio of signal turns squared to total resistance (N^2/R) approximately equal to 1,000.

To simulate the magnetic-amplifier operation, the voltage wave form applied to the core must be as close as possible to that which occurs in the actual amplifier.

In addition to fulfilling these operating requirements, the problem of instrumentation is important. The accuracy of the core test data depends on the accuracy of the instrumentation regardless of how well the circuit simulates low-impedance operation.

The low-impedance test method and circuit described in this paper fulfill all of these requirements in the laboratory.

TESTER OPERATION

The low-impedance core-tester circuit diagram is shown in Fig. 10(A). This tester applies a voltage across the test core with a wave form as shown in Fig. 11. The voltage wave form is sinusoidal until the firing angle at which time the applied voltage drops to a negligible value. The firing angle of the applied voltage is controlled by controlling the firing angle of the thyratron tube V_1.

The basic principle of operation of the circuit is as follows: A sinusoidal voltage of a specific value is applied across a specified number of primary turns of saturating transformer T_2. The voltage

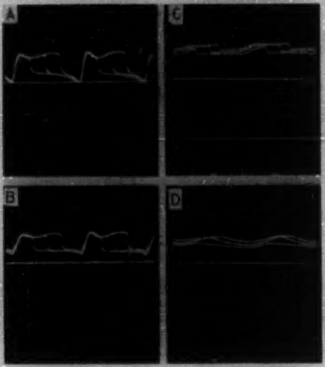


Fig. 4 (left). Reset current wave forms for varying control impedance operation of Deltamax

A—Low, $N^2/R=1{,}000$
B—Intermediate, $N^2/R=160$
C—Approaching high, $N^2/R=10$
D—High, N^2/R approaching zero

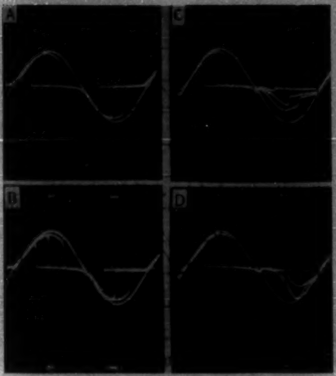


Fig. 5 (right). Gate-winding voltage wave forms for varying control impedance operation of Deltamax

A, B, C, D same as in Fig. 4, respectively

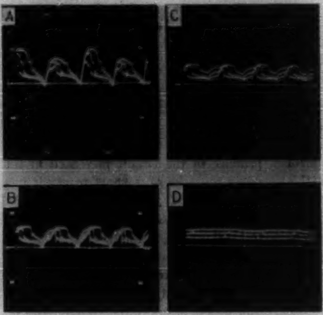


Fig. 6 (left). Reset current wave forms for varying control impedance operation of Supermalloy

A, B, C, D same as in Fig. 4, respectively

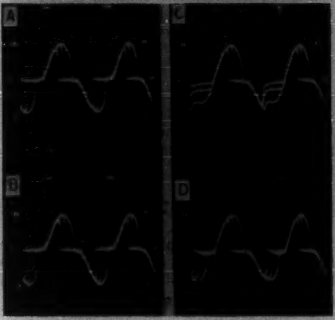


Fig. 7 (right). Core voltage wave form for varying control impedance operation of Supermalloy

A, B, C, D same as in Fig. 4, respectively

Vol. 7
Pt. 1
1956

is transformed and applied across the turns of the test core, but retains its sinusoidal wave shape. When the thyratron V_1 fires on the first half-cycle or reset, the applied voltage drops to the voltage drop across the thyratron or approximately zero since the ratio of supply voltage to tube drop is made high in the circuit. During the second half-cycle or set, the transformer T_2 will saturate itself at the same point in the half-cycle. Both firing angles occur at the same respective point because the positive volt-time area across a transformer must equal the negative volt-time area. By controlling the volt-time area of one half-cycle, the volt-time area of the second half-cycle is automatically controlled. The resulting applied wave form across the core is approximately symmetrical and the same as would occur across the core in a low-impedance control magnetic amplifier as shown in Figs. 12(A) and (B). These voltage wave forms were measured directly across the core winding of the amplifier and the turns of the test connector. The reset voltage is the negative half-cycle.

The test core does not have to be driven into saturation to operate on its minor loop as it does in a magnetic amplifier. An inherent unbalance in the circuit causes the core always to seek saturation in one direction as is shown in Fig. 14(C). This will be explained later in the detailed description of the circuit operation.

Only the square-loop material Deltamax was investigated. Core materials with *B-H* loop characteristics that differ from the square-loop properties of Deltamax may have to be driven into saturation.

The actual control characteristic of the core, which is a measure of the minor loop characteristics of the core, can be measured with the circuit. The amount of reset of the core from saturation is controlled by the firing angle of the thyratron. If the thyratron fires immediately at the beginning of the cycle, the core will not reset and will remain in saturation simulating the full "on" position of the magnetic amplifier. If the thyratron does not fire at all, the core flux will reset for the entire half-cycle corresponding to the cutoff position of the magnetic amplifier. The maximum amount of reset is determined by the specified voltage transformed by T_2. These are the extremities of the control characteristic.

Correlation Between Tester and Amplifier Data

To determine the ability of the low-impedance core tester to measure the operating characteristics of the toroidal core under low-impedance control conditions, data were taken on the core tester and compared with data taken on the full-wave center-tapped circuit with a low-impedance control. The reactor was made up of two Deltamax cores (1.0-inch ID by 1.375-inch OD by 0.25-inch 2-mil tape).

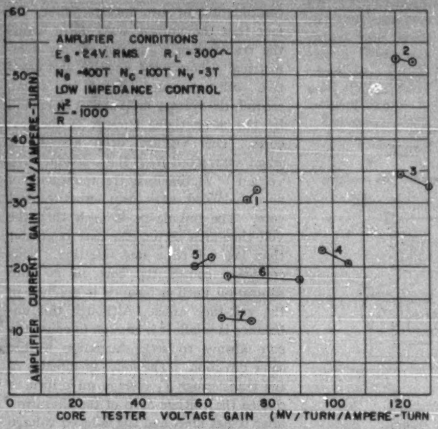


Fig. 8. Comparison of amplifier current gain versus constant-current reset core tester voltage gain for seven matched pairs of Deltamax cores

Grain-oriented 50% nickel-iron 2-mil tape. 1.0 inch ID by 1.375 inch OD by 0.25 inch

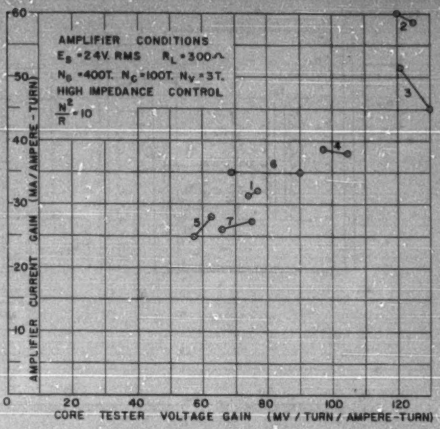


Fig. 9. Comparison of amplifier current gain versus constant-current reset core tester voltage gain for seven matched pairs of Deltamax cores

Seven matched pairs of the Deltamax cores were selected for the tests, with *B-H* loop widths that varied considerably. This was done to produce as wide a spread of characteristics as possible so as to present a more realistic picture of the resulting correlation.

The control characteristic voltage gains of the cores were first measured on the low-impedance core tester as will be explained later. The amplifier current gains were then measured by use of the full-wave center-tapped circuit after each pair of cores was wound into a reactor. Care was taken to measure the gains while operating the core under the same conditions in the tester and the amplifier. Care was also taken to measure the gain over the same region of the control characteristic in both cases. This would eliminate any discrepancies in the data due to nonlinearities in the control characteristic. This same procedure was used to take the data of Figs. 8 and 9.

The amplifier current gain was plotted versus the tester voltage gain and the resulting correlation is shown in Fig. 13. Perfect correlation would result if all the points fell on a straight line passing through zero. Only a representative figure could be used as a measure of the voltage gain of the tester because of the method used to measure the reset current which will be explained later.

The ability of the low-impedance core tester to measure the operating characteristics of a core can be appreciated more by a comparison of the results of Fig. 13 with those of Fig. 8.

Comparison of Tester Operation With Amplifier

The ability of the low-impedance core-test circuit to simulate the actual magnetic amplifier operation under low-impedance control conditions can be shown more clearly by comparing oscillograms of voltage and current wave shapes.

The oscillograms presented are of magnetizing current wave forms, pickoff voltage wave forms, and minor *B-H* loops, and compare the core tester operation with that of the full-wave center-tapped circuit under identical operating conditions.

The magnetizing current wave forms showing the core reset current, which is the portion above the zero axis, are presented in Figs. 14(A) and (B); they show the same core in the core tester and in the amplifier, respectively. A minor *B-H* loop characteristic of a different core in the core tester and in the amplifier, with the same firing angle as in Figs. 14(A) and (B), is shown in Figs. 14(C) and (D).

A comparison of Figs. 14(A) and (B) and 14(C) and (D) shows conclusively that the low-impedance tester closely simulates the actual operation of the magnetic amplifier under low-impedance control conditions.

The reset current portion of the wave forms shown in Figs. 14(A) and (B) differs slightly after the firing angle because of a coupling between the core windings of the magnetic amplifier in the circuit used. In the full-wave center-tapped circuit, when one core comes out of saturation a voltage is induced in the winding of the other core, which is being reset. This causes the reset current to slope towards zero,[3] and occurs in the region after the reset current drops suddenly to the smaller value. The current in Fig. 14(B) slopes towards zero whereas the current in Fig. 14(A) remains horizontal due to the absence of this effect. The magnitude of this effect is small in square-loop materials due to the low slope of the *B-H* loop in saturation.

The magnetizing current wave forms also differ in the region where the core comes out of saturation. In the magnetic amplifier the core is driven far out into saturation whereas in the core tester it is not. This effect is eliminated by use of a mechanical chopper to chop off a small portion of the beginning of the reset half-cycle of current as shown in Fig. 15(B). This elimination is necessary as the slope of that portion where the current goes from the set half-cycle to the reset half-cycle changes with the firing angle because the core is not driven far into saturation as in the actual magnetic amplifier. To eliminate this variable, the chopper is used.

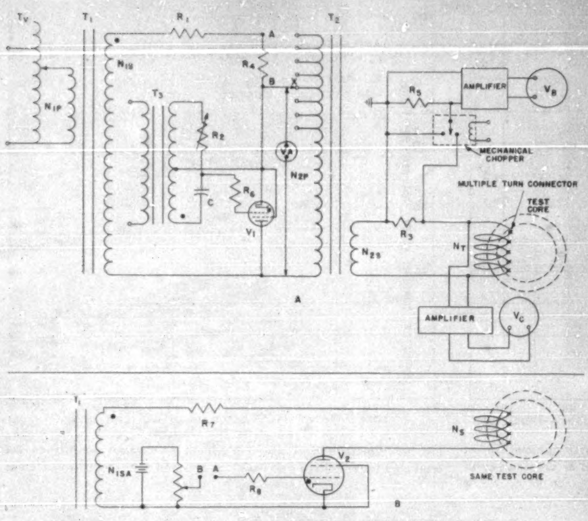


Fig. 10

A—Low-impedance core-tester circuit B—Circuit to drive core into saturation

The chopper is also used to take out the setting half-cycle of the magnetizing current so that only the reset half-cycle of current is metered. Fig. 15(A) shows the current wave form after it has been chopped to be metered. The chopper can perform both of the afore-mentioned functions because it can be made to be effective for only about 150 degrees of the half-cycle.

DETAILED DESCRIPTION OF CIRCUIT OPERATION

The diagram of the circuit to be described in the following is given in Fig. 10.

To test a given core, the core is placed in the loop of the multiple-turn connector and the connection made, closing the path of the test turns N_T about the core. With the thyratron V_1 turned off, the proper tap X of primary winding N_{2p} of T_2 is chosen and variac T_v varied until voltmeter V_A reads a specific prescribed voltage for the size and type of core to be tested. This primary tap and value of specified voltage are chosen so that the voltage applied across the test-core windings N_T will be the same percentage of the core saturation voltage as will be applied to the core in the actual amplifier circuit. This will apply a full sine wave of voltage to the core and simulate the "off" condition of the core.

The firing angle of the thyratron V_1 is varied by varying R_2, and thus the reset of saturable transformer T_2 is controlled. Since the volt-time area of the positive and negative half-cycles of the transformer must be equal, a symmetrical wave shape of voltage, as shown in Fig. 11(A), is applied to the test core. By varying the firing angle of V_1, the firing angle (a) of saturable transformer T_2 is varied. This wave shape is very similar to the one applied across the winding of the core in the actual magnetic amplifier.

As mentioned previously, the test core always seeks saturation in one direction due to an inherent unbalance in the circuit. The unbalance is caused by the combination of two factors: first, the fact that the voltage wave form across the transformer T_2 is not exactly symmetrical although the positive and negative volt-time areas must be equal; second, the voltage drop across R_3 caused by the magnetizing current of the test cores. This voltage drop across R_3, when subtracted from the secondary N_{2s} voltage of T_2, results in the application of an unbalanced wave form across the test core. The unbalance is such that the volt-time area setting the core is greater than the volt-time area resetting. This condition causes the core to go into saturation until a balance is reached in the volt-time areas. Although this unbalance is small, it is enough to cause the core always to seek saturation in the same direction. The small asymmetry of the transformer T_2 voltage wave form is due to the voltage drop of the thyratron after it has fired and the induced voltage by the transformer as it comes out of saturation. Both of these voltages are of the same polarity as can be seen in Fig. 12(A). These voltages are small and do not affect the firing angles of the applied wave form appreciably but they are enough to cause this desirable small unbalance.

To obtain characteristic data on the core being tested, the magnetizing current of the core and the core-induced voltage are measured. The reset portion of the core magnetizing current is measured by metering the drop across R_3. The reset half-cycle of current only, is measured by a mechanical chopper. The reset half-cycle drop across R_3 is then amplified and measured by V_B. The meter can be calibrated to read the equivalent magnetizing current.

The induced voltage is measured directly across the turns on the test core or across a separate pickoff source of three turns on the connector to eliminate any slight inaccuracy due to the very small voltage drop caused by the resistance of turns N_T. The core voltage is meas-

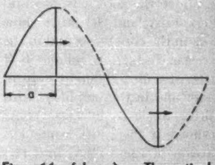
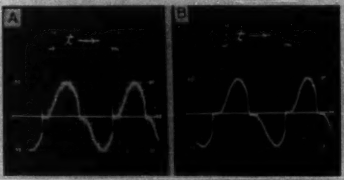


Fig. 11 (above). Theoretical tester voltage wave form



A—Tester B—Amplifier

Fig. 12 (right). Comparison of the voltage wave form applied across the core between the actual amplifier and the low-impedance tester

Control impedance ratio $N^2/R = 800$ for both ured by a meter that amplifies voltage and measures its average value.

From these current and voltage measurements, a complete gain curve for the core may be plotted which will predict the core's characteristics in the actual magnetic-amplifier circuit.

The resistance of R_3 may be varied to simulate any practical level of control impedance desired although the circuit is primarily designed for low-impedance testing.

Step-up transfomer T_1 is used to raise the supply voltage level so as to minimize the effect of the voltage drop across thyratron V_1 when it fires. Resistor R_1 is made only large enough to limit the forward current through V_1 when it is conducting.

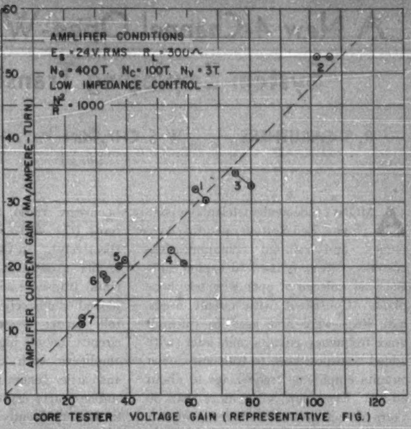
The ratio of the test-core turns squared N_T^2 to the total resistance around the paths from N_2, to N_T must be kept extremely low to insure that the voltage applied to the core is sinusoidal. This also applies to the primary circuit of T_1. Care must be taken to keep the contact resistance low in the connector.

If it becomes necessary or desirable to drive the test core into saturation, circuit *10B* may be used.

The polarity of N_{1-4} of T_1 is such that when the saturable transformer T_2 goes into saturation, the drop across R_4, which is connected to points *A* and *B* of circuit *B*, will fire thyratron V_2. The resulting ampere-turns through the few turns of N_3 about the test core will drive the test core into saturation.

It should be noted that the response time of the thyratron is important. For example, to test cores at 400 cps, the thyratron must operate satisfactorily at that frequency.

Fig. 13. Comparison of amplifier current gain versus low-impedance core-tester voltage gain for seven matched pairs of Deltamax cores



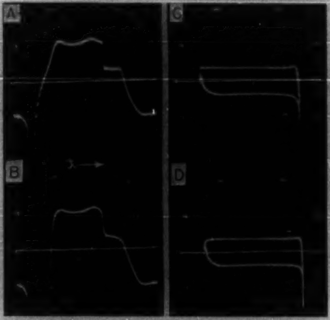


Fig. 14 (above). Comparisons of reset current and B-H loop operation between the actual amplifier with low-impedance tester

A—Tester magnetizing current
B—Amplifier magnetizing current
C—Tester B-H loop
D—Amplifier B-H loop

Fig. 15 (above). Core magnetizing current in low-impedance tester before and after going through mechanical chopper to be measured

A—Before B—After being chopped

Conclusion

The results presented in this paper show that this method of testing toroidal cores closely simulates the low-impedance control operation of the actual magnetic amplifier. The good correlation resulting between the gain characteristics measured on the test method and the actual amplifier operating under low-impedance conditions shows that the test method can determine the low-impedance operational characteristics of a core without the necessity for winding a large number of turns around the toroidal core.

References

1. Magnetic Characteristics Pertinent to the Operation of Cores in Self-Saturating Magnetic Amplifiers, R. W. Roberts. *AIEE Transactions*, vol. 73, pt. I, 1954, pp. 682–90.

2. Magnetic Amplifier Gapless Core Tests, J. R. Conrath. *Electronics*, New York, N. Y., vol. 25, Nov. 1952, pp. 119–21.

3. Magnetic Amplifiers (book), H. F. Storm et al. John Wiley and Sons, Inc., New York, N. Y., 1955, pp. 236, 262–96.

A New 4-Channel Open-Wire Carrier System Employing Transistors

R. S. CARUTHERS
MEMBER AIEE

W. S. CHASKIN
ASSOCIATE MEMBER AIEE

H. K. KRENGEL
NONMEMBER AIEE

AMONG independent telephone companies, Bell System operating companies, and railroad communications systems, carrier is used to exploit fully the vast mileage of open-wire telephone plant. Consistent with circuit needs, wire lines were first used to transmit voice-frequency circuits and, later, with added transpositions, to transmit carrier circuits employing frequencies to about 35 kc. Even now, after more than 30 years of such growth, however, the great majority of open-wire message traffic for short- and medium-haul distances is transmitted at carrier frequencies below 35 kc. The number of circuit miles in these carrier frequencies is now in the millions. Continued major growth is expected in open-wire plant, not by adding more wire lines but by moving to higher frequencies.

Past system design practices have generally advocated sufficient transpositions of wire lines to make crosstalk tolerable at the top frequencies of the carrier systems involved. In more recent years, however, the availability of inexpensive compandors has permitted wire lines to be used at frequencies much beyond those previously considered satisfactory from a crosstalk point of view. In the case of 35-kc transposed lines, compandors make the lines suitable for frequencies up to about 80 kc, and thus allow the transmission of four additional carrier channels. Full use of these additional channels will nearly double present open-wire circuit mileage.

This paper describes a 4-channel open-wire carrier system, the Lenkurt type *45CB*, for use in the 40- to 76-kc spectrum above present carrier systems using frequencies up to 35 kc. Channel terminal equipment, including compandors, makes use of miniature equipment already used in other Lenkurt open-wire, radio, and cable carrier systems (the *45A*, *45BX*, and *45BN* respectively). In the common amplifying, carrier generating, and regulating circuits, transistor circuitry has been employed extensively for the first time in a toll carrier system. The transistor circuits include high-gain high-frequency amplifiers with considerable feedback and carrier-generating circuits with highly stable quartz crystals. Design of these transistor circuits has been a highly satisfactory experience, and performance results are excellent. A detailed discussion of the design of transistor circuits for the system is also given.

Electrical Features

As in all types of Lenkurt *45*-class carrier, the first stage of modulation translates the speech frequencies of each message channel into a carrier band in the frequency region from 8 to 24 kc. In the *45CB* system, as indicated in Fig. 1, four channels are derived by modulating carriers of 12 and 20 kc. Channels 1 and 2 are derived from the lower and upper side bands of the 12-kc carrier, and channels 3 and 4 are derived from the lower and upper side bands respectively of the 20-kc carrier. This system of modulation is generally referred to as twin-channel single-side-band operation. The 4-channel band from 8 to 24 kc is further translated to line frequencies in either a 40- to 56-kc band or a 60- to 76-kc band. Where a terminal is transmitting the low-group band, it is receiving the high-group band. This arrangement is termed a low-group transmitting terminal. An alternate high-group transmitting terminal transmits a high group and receives a low group.

In a system without repeaters, a high-group transmitting terminal is required at one end and a low-group transmitting terminal at the other. In repeatered operation, either type may be required at either terminal depending upon the particular system layout. Each repeater transposes low and high groups so that transmission over two or more sections is alternately in high- and low-frequency bands. Such repeaters are commonly called "frequency-frogging repeaters."

In traversing a frequency-frogging repeater, the frequency order of the channels in a group is reversed. This requires that the frequency order of the high and low groups be opposite, so that at end terminals the received group can be properly demodulated. To accomplish the reversal, the low group is formed from the lower side band of a 64-kc carrier and the high group is formed from the upper side band of a 52-kc carrier.

In Fig. 2, a block diagram of the terminal is shown with message and signal levels indicated. To the left are shown the standard compandor and channel equipment used in all *45*-class Lenkurt equipment for open-wire, cable, and radio systems. The wire frequency terminations are arranged for 2- or 4-wire voice-frequency operation at all of the standard telephone transmission levels, which allows either single- or multilink operation in the toll-switching telephone plant.

The equipment common to the four channels is shown to the right and consists of:

1. A transmitting unit which provides amplification and translation between the channel frequencies at 8 to 24 kc and the line frequencies at 40 to 56 kc, or 60 to 76 kc.
2. A receiving unit which provides amplification, regulation, and conversion between line frequencies and the receiving 8- to 24-kc channel band.
3. A combining directional filter unit for transmitting and receiving groups.
4. A carrier supply unit for deriving the group and channel frequencies from two crystal oscillators at 32 and 52 kc.

REGULATION PERFORMANCE

In providing gain and regulation, a common practice is to engineer performance for three sets of line conditions: dry weather, wet weather, and sleet. The amount of sleet designed for is near the maximum the wires can carry without breakage. This is normally considered to be about 1/4 inch of radial-diameter ice and results in about 50 decibels (db) of attenuation for a 70-mile section at the top frequency of the *45CB* system.

In Fig. 3, attenuation characteristics of typical line sections are shown for various weather conditions. Included are curves for 50 miles of lines with 1/4 inch of ice, 100 miles of wet-weather line, and 100 miles of dry-weather line. Also shown are attenuation characteristics for tandem sections with an intervening frequency-frogging repeater. The benefits of frequency frogging in all weather

---

Paper 56-245, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 21, 1955; made available for printing December 6, 1955.

R. S. CARUTHERS, W. S. CHASKIN, and H. K. KRENGEL are with the Lenkurt Electric Company, Inc., San Carlos, Calif.

conditions in eliminating slope characteristics are apparent. Bulge across a single channel is so slight that it can be ignored. Because of the small residual slope and bulge, the flat types of regulators provided by repeater and terminal equipment are adequate. Channel regulation supplementing the group regulation corrects the small remainder of level inequality among individual channels.

The adequacy of regulation provided by group- and channel-regulating techniques can be seen from the curves in Fig. 4. These curves show that changes in message output level are so slight through extremes of line loss that satisfactory performance can be expected under most conditions of plant layout. Fig. 5 shows the channel frequency characteristic.

NOISE PERFORMANCE

Design of the system for adequate noise performance entails proper distribution of noise contributions between terminal and line effects during both normal and storm operating periods. During normal periods, most of the noise disturbances to channels result either from various modulation products of signal tones, or from crosstalk disturbances caused by intermodulation products in common amplifiers and modulators. In general, noise performance during normal periods is better than required to allow for the major increases in noise disturbances during adverse weather conditions.

During abnormal weather with severe localized thunderstorms, either signal or message channels, or both, may be seriously affected by atmospheric noise. In Fig. 6, noise measurements are shown observed during field tests. These data are undoubtedly pessimistic, as it was necessary to increase the line attenuation during these heavy, localized storm periods to simulate continuous wet weather over the entire line. It appears doubtful that localized lightning storms will be coincident with steady rain over the whole line section. In field tests, very little increase in attenuation has been observed during the severe local thunderstorm period.

During icing conditions, lines are invariably quite free from high atmospheric noise. Only the terminal and repeater equipment contribute to noise during such periods. However, except for excessive attenuations with heavy icing, where the regulation systems run beyond their ranges, the signaling and speech transmission is nearly as satisfactory as in normal weather periods.

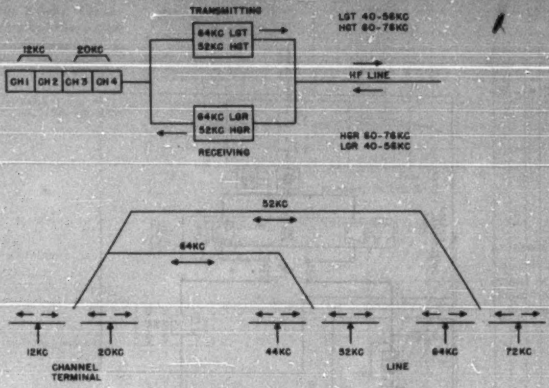


Fig. 1. Modulating plan and frequency allocation

CROSSTALK PERFORMANCE

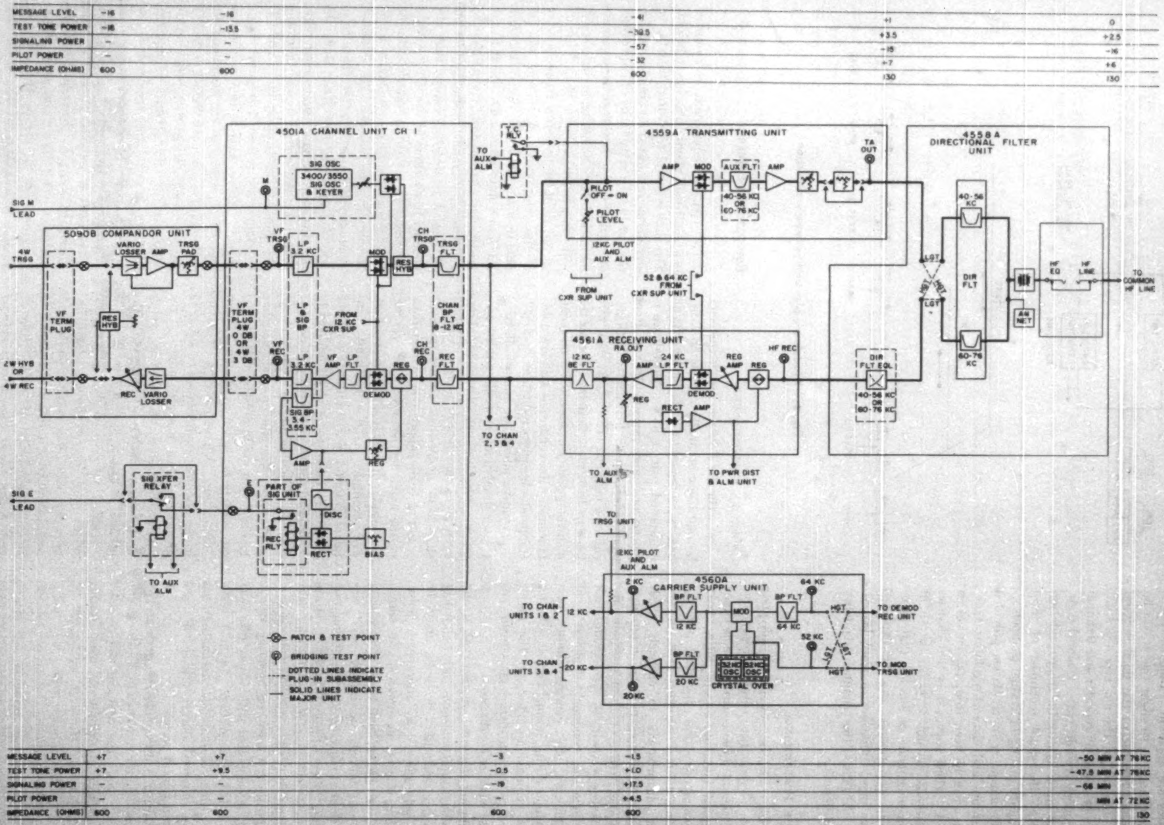
Since the *45CB* system uses frequencies beyond those originally considered satisfactory, in regard to crosstalk for a 35-kc transposed pair, a major dependence is placed on improvements derived from compandor action. Frequently, both far-end and near-end crosstalk in such wire lines are little more than 30 db below the disturbing source at an equal-level point. During dry-line conditions with low noise, compandor benefit of as little as 23 db (28-db expandor advantage minus 5-db compressor disadvantage) would not be adequate. The Lenkurt compandor, however, gives considerably more advantage than 23 db under these circumstances.

In Fig. 7, the suppression characteristic of the expandor is shown for varying magnitudes of noise and speech crosstalk. For example, a 0-volume-unit (vu) talker (about 1 in 500 talkers) would be effectively raised to +2.5 db referred to 1 milliwatt (dbm) speech power because of compressor action at the zero message level point. The line crosstalk coupling loss of 30 db would present a disturbing crosstalk power to the expandor of about −27.5 dbm. Approximately a 33-db loss in the expandor would reduce crosstalk to about −60 dbm of speech power at the expandor output at the zero level point, or to approximately 60 db below the original disturbing speech.

Table I shows crosstalk expectancy in relation to talker volume. Equal-level crosstalk of 30 db is assumed on the lines and, arbitrarily, a talker is assumed to be talking one-fourth of the time from one end of the circuit (25-per-cent activity factor).

Equipment Features

The terminal equipment is mounted in a small shelf about 16 inches in height. This shelf is completely wired, with fusing and alarms integrated for simple installation on 19-inch relay racks of either duct or channel type. Standard arrangments for back-to-back mounting are employed, with projections no more than 5 inches to the front or in depth from the face of the relay rack. A bottom row of plug-in units, shown in Fig. 8, contains the four compandor units. Next above are the four channel units. Both channel and compandor units are standard for the 45 line of equipment, except that the channel band filters of the *45CB* system are designed for twin-channel operation and take advantage of compandor action to simplify the design. The common equipment units on the top shelf are, from left to right: the fuse alarm unit, the directional filter unit, the receiver regulator unit, the carrier supply unit, and the transmitting amplifier unit. Only 21 vacuum tubes are used in the entire equipment. One of these is in the common equipment, where as yet techniques for wide-range regulation have not permitted substitution of a transistor for the vacuum tube. As in other *45*-type equipment, only long-life miniature triodes and pentodes are used. These tubes have 20-volt heaters permitting operation from 24- or 48-volt office batteries as well as from alternating current. The need for a 130-volt battery for the vacuum tubes makes it economical to design the common-equipment transistor circuitry on a standard 130-volt basis to avoid optional strap-



MESSAGE LEVEL
TEST TONE POWER
SIGNALING POWER
PILOT POWER
IMPEDANCE (OHMS)
-16 -16 — — 600
-16 -13.5 — — 600
-41 -39.5 -57 -32 600
+1 +3.5 -15 +7 130
0 +2.5 -16 +6 130
4501A CHANNEL UNIT CH 1
SIG OSC
3400/3550 SIG OSC & KEYER
M
SIG M LEAD
5090B COMPANDOR UNIT
VARIO LOSSER
AMP
TRSG PAD
4W TRSG
VF TERM PLUG
RES HYB
2W HYB OR 4W REC
REC
VARIO LOSSER
VF TERM PLUG 4W 0 DB OR 4W 3 DB
VF TRSG
LP 3.2 KC
LP & SIG BP
VF REC
LP 3.2 KC
SIG BP 3.4 - 3.55 KC
VF AMP
LP FLT
MOD
RES HYB
FROM 12 KC CXR SUP
DEMOD
REG
CH TRSG
TRSG FLT
CHAN BP FLT (8-12 KC)
CH REC
REC FLT
AMP
REG
SIG XFER RELAY
SIG E LEAD
TO AUX ALM
E
PART OF SIG UNIT
REC RLY
DISC
RECT
BIAS
T.C. RLY
TO AUX ALM
TO CHAN 2, 3 & 4
4559A TRANSMITTING UNIT
PILOT OFF = ON
PILOT LEVEL
12KC PILOT AND AUX ALM
FROM CXR SUP UNIT
AMP
MOD
AUX FLT
40-56 KC OR 60-76 KC
AMP
TA OUT
52 & 64 KC FROM CXR SUP UNIT
4561A RECEIVING UNIT
12 KC BE FLT
RA OUT
AMP
24 KC LP FLT
DEMOD
REG AMP
REG
HF REC
REG
RECT
AMP
TO AUX ALM
TO PWR DIST & ALM UNIT
4558A DIRECTIONAL FILTER UNIT
40-56 KC
DIR FLT
60-76 KC
LGT
HGT
DIR FLT EQL
40-56 KC OR 60-76 KC
AN NET
HF EQ
HF LINE
TO COMMON HF LINE
TO TRSG UNIT
12KC PILOT AND AUX ALM
4560A CARRIER SUPPLY UNIT
2 KC
BP FLT
12 KC
MOD
BP FLT
64 KC
64 KC
52 KC
HGT
LGT
TO DEMOD REC UNIT
TO MOD TRSG UNIT
TO CHAN UNITS 1 & 2
12 KC
TO CHAN UNITS 3 & 4
20 KC
BP FLT
20 KC
20 KC
52 KC OSC
64 KC OSC
CRYSTAL OVEN
PATCH & TEST POINT
BRIDGING TEST POINT
DOTTED LINES INDICATE PLUG-IN SUBASSEMBLY
SOLID LINES INDICATE MAJOR UNIT
MESSAGE LEVEL
TEST TONE POWER
SIGNALING POWER
PILOT POWER
IMPEDANCE (OHMS)
+7 +7 — — 600
+7 +9.5 — — 600
-3 -0.5 -19 — 600
-1.5 +1.0 +17.5 +4.5 600
-50 MIN AT 76 KC
-47.5 MIN AT 76 KC
-66 MIN
MIN AT 72 KC
130

Fig. 2. Block and level diagram

Vol. 7
Pt. 1
1956

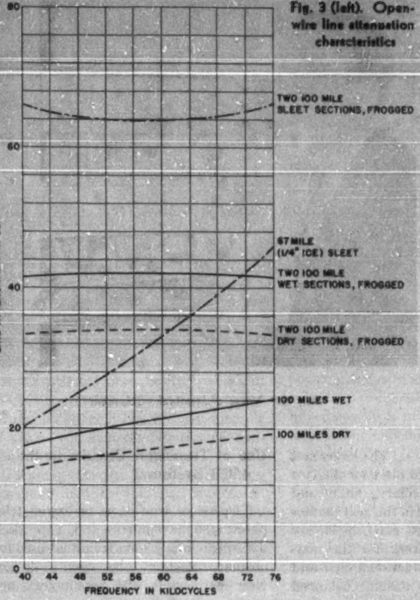


Fig. 3 (left). Open-wire line attenuation characteristics

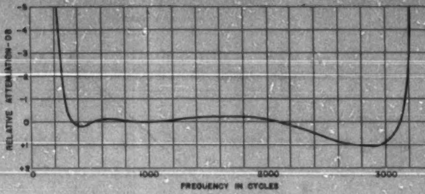


Fig. 5 (above). Channel frequency characteristic

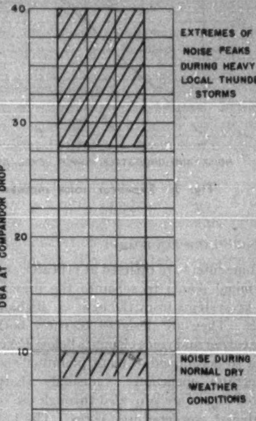


Fig. 6 (right). Channel noise during normal and storm periods

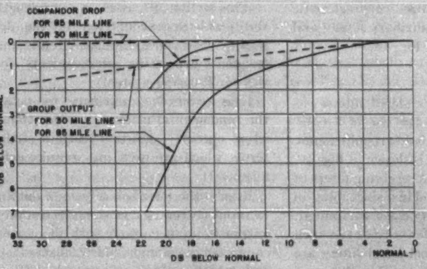


Fig. 4. System regulation performance

ping for tube filament battery conditions encountered in various telephone offices.

The total power drain is as follows: For the 48-volt filter battery, 1.01 amperes, and for the 130-volt plate battery, 0.244 ampere.

In Fig. 9, the equipment is shown with plug-in units removed. Twenty-pin gold-plated jacks are shown, with free-floating pins to accommodate similarly gold-plated plugs on the plug-in units. The fuse alarm unit is shown at the upper left, accessible by a hinged-cable arrangement. As in all *45*-type equipment, the locking and camming action of the latch on the individual plug-in units allows simple removal and insertion.

Alarms indicate the fuse or system failure with lights and keys standard in modern toll-office practice. In addition, an auxiliary alarm arrangement used with dialing systems can be associated with the terminal shelf to provide automatic disconnect and automatic busying of signaling circuits in case of system failure. Also, the auxiliary alarm equipment enables operation to remote unattended terminals where a test can be made from the attended terminal as to clearance of line trouble; this is followed by reset of distant alarm relays.

All the common-equipment plug-in units are quite similar in internal construction. Typical is the transmitting unit shown in Fig. 10. Transistors are mounted internally on boards in the same manner as other pigtail components, with the exception that a transistor socket is used. It is expected that transistor sockets will not be required within a short period of time; in fact, they may not be needed now. However, pending more extensive data, transistors are being supplied with a socket to allow their replacement should troubles develop over the next few years. Another benefit of the socket is that, if superior transistors do become available, they may be inserted in the equipment without soldering.

An unusual feature of the carrier supply unit is the employment of a crystal oven with dual crystals at 32 and 52 kc mounted within the same oven. Oven-temperature control to ±1 degree assures high stability of the frequencies for the channel carriers as well as the group carriers.

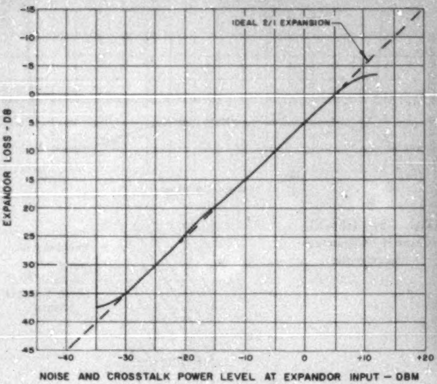


Fig. 7. Expandor noise suppression characteristic

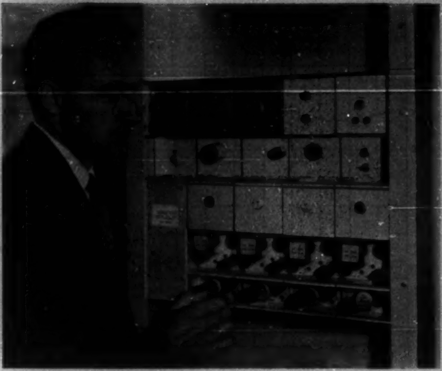


Fig. 8. View of terminal equipment

LINE FILTER EQUIPMENT

Line filters are injected at repeater and terminal points to separate the present 35-kc systems from the new *45CB* channels. They are designed to carry both the carrier and voice channels below 35 kc, and to accommodate both phantom and telegraph operation. Unusually sharp cutoff filters are provided through use of carbonyl powdered-iron cores. The low operating levels and the use of compandors in the system permit unusually small filters; both high-pass and low-pass line filters are contained in a single can of $7^{1}/_{2}$ by $3^{1}/_{2}$ by $3^{3}/_{16}$ inches. Two line filter sets for the two pairs of the phantom group can be contained within one pole-mounted filter housing, hung on the crossarm between the two pairs. A view of the pole-mounting line filter and housing is shown in Fig. 11.

The characteristic of the line filter is shown in Fig. 12. Reflection coefficient of the combined filter, directional filter, and terminal-repeater equipment is in the 10- to 20-per-cent range. Design was on the basis that, with compandors always required in the equipment, reflected near-end crosstalk will be at least 14 to 20 db below far-end crosstalk as contributed from each reflection point.

DIRECTIONAL FILTERS

Characteristics of directional filters are shown in Fig. 13. Both low-group and high-group directional filters are supplemented by an auxiliary directional filter operated at lower levels, with associated equalization for the combined directional filter complement and line filters. The mechanical arrangement of the directional filter is shown in Fig. 14. The larger can contains the directional filter for the two directions. The auxiliary filter and equalizer are contained in the next smaller can. The two smallest cans shown are for an annulling network for the nonflanked edges of the directional filter and for a repeater coil to provide balanced impedance to the high-frequency line. The combined equalized characteristic of directional filters, auxiliary filters, and line filters is shown in Fig. 15.

CHANNEL FILTERS

The design of the channel filters was simplified because of the use of a compandor to ease transmission requirements. A filter characteristic is shown in Fig. 16. Both transmitting and receiving filters of a channel are contained in a single plug-in unit, as with other *45*-class equipment. The channel units differ only as to which of four plug-in channel band filters are used. The equalization within the voice amplifier of the channel unit provides transmission channel characteristics comparable with those of other *45*-class equipment.

Use of Transistor Circuits in the 45CB System

The use of transistors in carrier telephone systems requires a design approach differing in many ways from that used for ordinary vacuum-tube circuit design. New parameters and terminology are necessary.

This section discusses at some length the problems encountered in the design of wide-band amplifier circuits and frequency-generating arrangements for the *45CB* carrier system.

It is assumed unnecessary to detail the fundamental manner in which transistors operate; however, certain of the terms which are used may require definition.

Briefly, the term α is a ratio of output to input current. β is a similar ratio of output-to-input current, with the circuit connected in common-emitter fashion. α and β cutoff are frequencies at which these respective current gain ratios are down 3 db. The term I_{c0} is the current flowing in the collector circuit with no bias between emitter and base.

Table I. Crosstalk Expectancy in Relation to Talker Volume

Talker Volume, Vu	Per Cent of Talkers	Per Cent Time Active	Approximate Compressed Power, Dbm	Power at Expandor Input,* Dbm	Power at Expandor Output, Dbm	Approximate Disturbance, Db Adjusted
0	0.2	0.05	+2.5	−27.5	−60	+22
−3	1	0.25	+1	−29	−63	+19
−6	4	1.0	−0.5	−30.5	−66	+16
−9	10	2.5	−2	−32	−68	+14
−12	25	6.25	−3.5	−33.5	−70	+12
−16	50	12.5	−5.5	−35.5	−72	+10

* 30-db crosstalk.

Fig. 9. Terminal equipment housing

Fig. 10. Transmitting group-equipment unit

Fig. 11. View of pole-mounting line filter and housing

TRANSISTOR CIRCUIT CONFIGURATIONS

Characteristics of the three commonly used transistor circuit configurations are shown in Table II. Space does not permit discussion of all of the circuits used in the *45CB* system. However, the transmitting amplifier of Fig. 17 is an example of a near-optimum combination of transistor arrangements.

Requirements for linearity, frequency response, load capabilities, impedance, and gain stability were quite rigorous. To achieve maximum power output with minimum distortion and good high-frequency response, the grounded-base configuration became mandatory in the output stage. Because of the low input impedance of this configuration, a common-collector (CC) interstage amplifier was used. An additional 30 db of required gain for the circuit made it desirable in the first stage of the amplifier to use the high-gain capabilities of the common-emitter (CE) circuit.

In discussing transistor circuits, it is convenient to make a comparison with familiar vacuum-tube circuits. However, because the input and output circuits are not isolated and because of the internal resistances of transistors, the internal feedback generated tends somewhat to modify and complicate circuit behavior. If the transistor is considered as a vacuum tube that always draws grid current and has resistances between all elements, then the analogy is valid.

The CE configuration is similar to the commonly used driven-grid circuit. Of the three configurations of Table II, it yields the highest power gains. Such an arrangement is shown in the input stage of the *45CB* transmitting amplifier; see Fig. 17. Gain in this circuit is expressed mathematically as in the following (assume generator is matched to transistor input).

$$\text{Power gain in db} = 20\ (\log\ \beta_2) + 10\ (\log\ R_L/R_I)$$

where

β_2 = current gain with load connected
R_L = load impedance
R_I = transistor input impedance

With transistors of present design, power gains of about 35 db can be obtained in practical CE circuits. Input impedance is a function of α and load impedance, and may be in the order of magnitude from 500 to 1,000 ohms. Output impedance is somewhat dependent on the internal parameters and the generator resistance R_G. A typical figure is 50 kilohms. High-frequency performance and distortion performance for the CE circuit are not as good as for the common-base (CB) configuration for the following reasons: in the CE connection, the upper frequency limit is $\beta_{\text{cutoff}} = \alpha_{\text{cutoff}}\ (1-\alpha)$, and directly α_{cutoff} in the CB arrangement. Since α is usually between about 0.95 and 0.99, the current gain will drop off much more rapidly for the CE connection.

Fig. 18 shows the collector-current characteristics V_c/I_c for the CE and CB configurations using a silicon transistor. These curves show that for minimum distortion at maximum power output, the grounded-emitter configuration is poorer than the grounded-base circuit.

The CB configuration used in the output stage of the amplifier in Fig. 17 is not only best suited where optimum power and distortion characteristics are required but also where the transistor is a limiting factor for high-frequency response. Some of the disadvantages of the CB configuration are low input impedance (30 ohms is typical) and relatively low gain. Although gain equations for the CB configuration indicate theoretical possibilities as good as the CE configuration, in practice this is not so. To achieve maximum gain, the load resistances assume impractical proportions. Further, to realize the low-distortion capabilities, emitter distortion must be minimized in both direct and feedback signal paths by using a generator impedance considerably

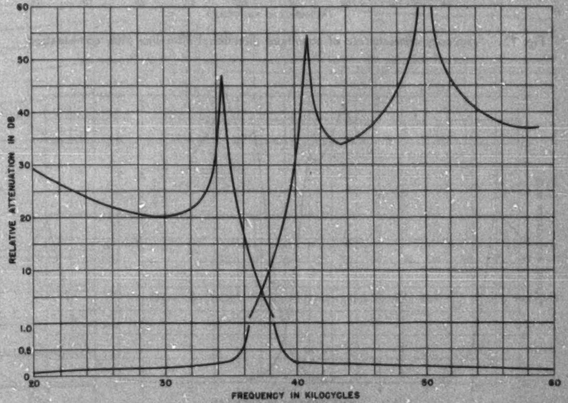


Fig. 12. Line filter attenuation characteristics

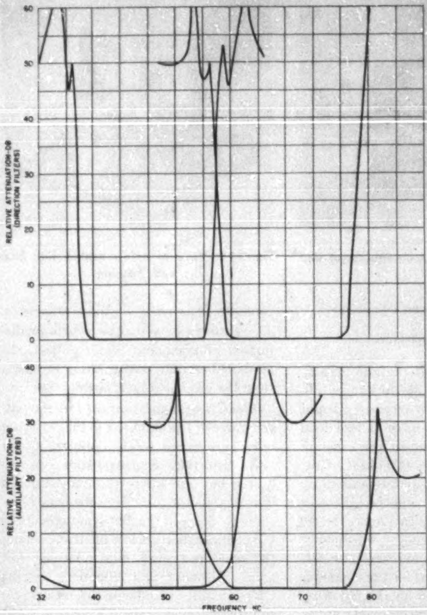


Fig. 13 (left). Directional filter attenuation characteristics

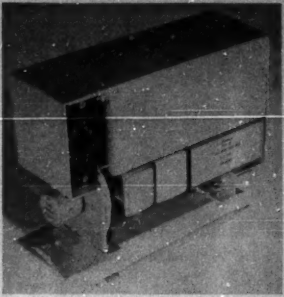


Fig. 14 (right). Plug-in directional filter unit

higher than the input impedance of the transistor. This causes a consequent loss in gain because of voltage division between the generator and the transistor. A typical gain figure is 20 db.

The CC circuit shown as the second stage of the amplifier in Fig. 17 is useful principally as an interstage device when connecting to a CB power stage or when operating into a low-impedance load. This circuit provides the lowest power gain of all three configurations (about 10 db). High-frequency and distortion characteristics are similar to the CE stage, except that maximum power output is limited by the permissible emitter voltage swing. For a particular transistor, the maximum power output is generally lowest for the CC configuration.

Temperature Considerations

The transistor is inherently a temperature-sensitive device. In germanium, the internal parameters vary considerably at rather moderate temperatures. In present germanium transistors, maximum operating temperatures occur somewhere between 50 and 70 degrees centigrade (C), allowing, of course, for a reasonable amount of power dissipation. Although the internal parameters for silicon also vary with temperature, the higher junction temperatures permissible (150 C) and a lower I_{c0} allow for sizable dissipation at ambients exceeding 100 C.

To achieve the circuit stability required in carrier design, compensation is necessary for temperature effects on both the d-c operating point and the a-c gain. In the amplifier shown in Fig. 17, this stability is attained by application of both a-c and d-c negative feedback. The a-c feedback circuits will be discussed in the following paragraphs. A typical d-c compensating network is formed by resistors R_1, R_2, and R_3 in Fig. 17. The transistor in this circuit is connected for

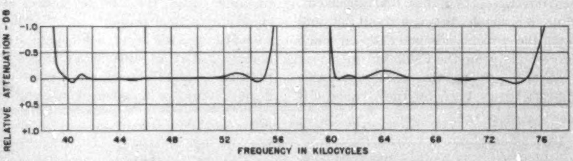


Fig. 15. Frequency characteristic of equalized directional filter–line filter combination

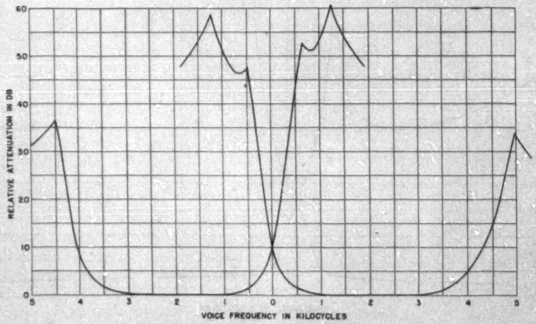


Fig. 16. Channel band filter attenuation characteristic

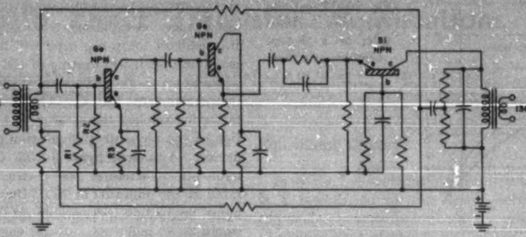


Fig. 17. Three-stage transmitting amplifier using transistors

Table II. Characteristics of Configurations

Circuit	Gain	Input Impedance	Output Impedance	High-Frequency Performance	Distortion	Tube Analogies
Common emitter (CE)	high	medium	medium	fair	poor	driven grid
Common base (CB)	medium	low	high	good	good	grounded grid
Common collector (CC)	low	high	low	like CE	fair	cathode follower

CE operation, which, as previously mentioned, is analogous to the driven-grid vacuum-tube circuit with a phase turnover of 180 degrees. Therefore, resistor R_3 is roughly equivalent to a cathode bias resistor and introduces stabilizing degeneration to the circuit. Because this resistor is by-passed by a rather large capacitor, the circuit is not degenerative at signal frequencies. Resistors R_1 and R_2 form a voltage-divider supplying battery potential between base and emitter to obtain the desired static operating currents. R_3 is made large compared with the emitter, base, and collector resistances and R_1 and R_2 are selected to provide a potential slightly higher than the voltage drop across R_3 at the proper emitter current.

The effectiveness of the temperature-compensating measures can be seen in Fig. 19, which shows gain and distortion for the amplifier. Note that gain is virtually constant for temperatures considerably in excess of 50 C, and that distortion variation is negligible.

FEEDBACK CIRCUITS

The entire amplifier of Fig. 17 is enclosed by a feedback loop. Although negative feedback is quite common in vacuum-tube circuits for carrier systems, the requirements are much more stringent when designing with transistors. Not only is the input impedance low as compared with vacuum tubes; but it is quite a variable from transistor to transistor and is subject somewhat to fluctuations of temperature and power supply. This, of course, is also true of the output impedance of the transistor. Furthermore, when designing high-frequency broadband equipment, the effect of the α or β cutoff and the attendant phase shifts must be carefully considered.

To achieve gain stability, flatness of frequency response, accuracy and stability of transmitting impedance, and low distortion, it is necessary to use relatively large combinations of both series and shunt feedback. The problem of matching the output stage to the load is similar to that encountered in a pentode tube. With the CB configuration, the output impedance is quite high, approaching one megohm; and for the CE configuration, is somewhat higher than the optimum load resistance. Therefore, it is necessary in the output stage to use a combination of series and shunt feedback (series raises the impedance; shunt reduces it), so proportioned that the desired net feedback is obtained with a matched-generator impedance. To obtain precision control of this impedance independent of transistor

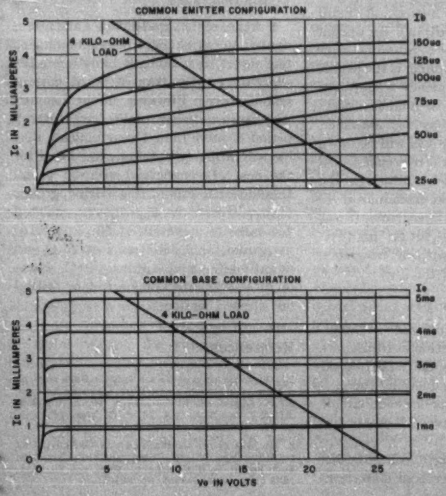


Fig. 18 (left). Static characteristics of CE and CB silicon transistor

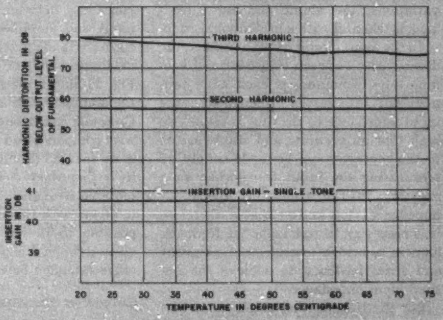


Fig. 19 (above). Effect of temperature on transistor amplifier gain and distortion characteristics

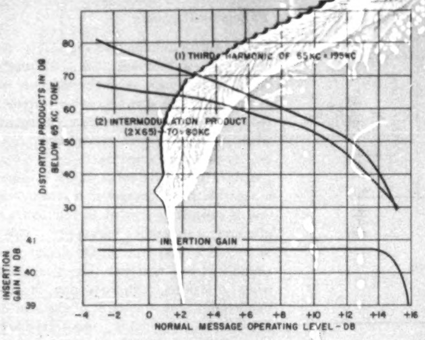
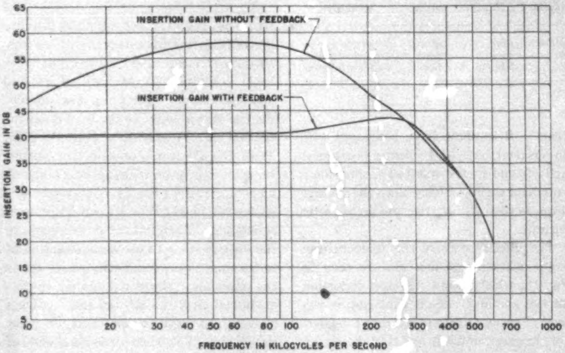


Fig. 20. Load and distortion characteristics of a 3-stage transistor amplifier

Fig. 21 (below). Frequency characteristic of a 3-stage transistor amplifier



variations, the shunt and series feedback is taken off with a bridge connection.

In the input circuit of the amplifier, the dissimilarity between transistor and vacuum-tube techniques is even more pronounced. At carrier frequencies, the input impedance to the tube is usually so high that an accurate and stable match to the generator can be made simply by terminating the input transformer in a resistor. In the transistor, the input impedance is low and a variable; therefore it is necessary to terminate the feedback loop in a bridge arrangement of shunt and series feedback to achieve the desired input impedance.

For the transmitting amplifier of Fig. 17, the input and output impedances are constant within ±5 per cent for any statistical distribution of transistors within the manufacturer's specification and over the entire termperature range from -20 to $+60$ C. Fig. 20 shows the load characteristics and some of the distortion products for this amplifier. Since the amplifier is operated at zero message level (zero loss between a 2-wire toll-circuit input and amplifier output), its output capacity is adequate for tone and message loading to which the amplifier is subjected. Fig. 21 shows the frequency response of the amplifier with and without feedback. Crosstalk contributed by the amplifier, with compandor advantage, will be down more than 65 db below a reference signal in a disturbed channel. Although the required frequency band extends from 40 to 76 kc, with the aid of feedback, it was possible to design for future high-frequency applications with a frequency range extended beyond 150 kc.

CHOICE OF TRANSISTORS

The carrier design engineer has available a rather large selection of transistors of the germanium types, and a fairly good selection in silicon. In choosing the particular type of transistor for each circuit of the *45CB* system, several factors were considered.

In power output stages, collector dissipation, ambient temperature, distortion, and frequency characteristics are of prime importance. A typical germanium power transistor of the p-n-p type will have a maximum dissipation rating of 150 milliwatts at 25 C, and a maximum junction temperature of 100 C. The derating factor for ambient temperature is a function of the physical configuration of the device. A typical figure is 0.5 C per milliwatt. Therefore, for a transistor rated at 150 milliwatt at 25 C, the maximum dissipation permissible at 60 C would be 80 milliwatts. Good engineering practice indicates that the transistor should be operated at approximately 50 per cent of rated dissipation at 25 C, providing maximum junction temperature is not exceeded at the extremes of operating ambient temperature. The output stage of the amplifier, shown in Fig. 17, required a dissipation of 80 milliwatts at 60 C, with an α cutoff beyond 1 megacycle. These requirements can be met at present only with silicon transistors. For driver stages, several germanium types are available. Although both p-n-p and n-p-n small-signal transistors can be obtained with α cutoffs considerably in excess of 1 megacycle, the p-n-p types tend to have considerably greater collector capacities, and, therefore, are not as well adapted for use in broad-band high-frequency amplifiers. Use of p-n-p transistors is therefore limited to amplifiers in the audio- or low-carrier-frequency spectrum, although they can be used quite successfully in oscillators to frequencies of 100 kc or higher.

LIFE EXPECTANCY

Because of the rapid advance and attendant frequent changes in manufacturing techniques, it is difficult to estimate life expectancy with transistors of current manufacture. However, in the Lenkurt laboratories, transistor life tests over a period of more than 1 year indicate that life will probably equal or exceed that for the best of commercial vacuum tubes. In additional tests at this writing, groups of transistors of five types have completed 4 months of operation at 50 per cent of maximum rated dissipation at 25 C and at an ambient temperature of 40 C. None of these transistors has failed or drifted out of specification.

References

1. APPLICATION AND TRANSMISSION FEATURES OF A NEW 12-CHANNEL OPEN-WIRE CARRIER SYSTEM, K. E. Appert, R. S. Caruthers, W. S. Chaskin. *AIEE Transactions*, vol. 73, pt. 1, Mar. 1954, pp. 18–27.

2. A MINIATURE COMPANDOR FOR GENERAL USE IN WIRE AND RADIO COMMUNICATION SYSTEMS, F. S. Boxall, R. S. Caruthers. *Ibid.*, vol. 72, pt. 1, 1953 (Jan. 1954 section), pp. 804–11.

Field Effect Transistor Applications

CHAANG HUANG
NONMEMBER AIEE

MELVIN MARSHALL
NONMEMBER AIEE

B. H. WHITE
NONMEMBER AIEE

Synopsis: A field effect transistor[1] is a unipolar device, distinguishable from a junction transistor in the following two ways: 1. the current flow is carried predominantly by one type of carrier alone; and 2. the electric field intensity is relatively high, thus the current flow is caused by field drift instead of by diffusion. Fig. 1(A) shows the construction of a field effect transistor. It consists of a piece of *n*-type germanium or silicon slab with a *p*-type junction formed around the center of the slab. Two ohmic contacts are placed in the vicinity of the junction. Three electrodes are connected to the transistor. The electrode connected to the junction is called the "gate." The electrodes connected to the ohmic contacts are called the "source" and the "drain." Five papers have already appeared on the field-effect transistors[1-5] in which the design theory, their experimental verification, the d-c equivalent circuit, and the estimate of the maximum cutoff frequency were presented. This paper describes an a-c equivalent circuit of the field effect transistor; the measured parameters of a type of germanium field effect transistor, and some of its circuit applications, with emphasis on oscillators and switching circuits. Special features of a field effect transistor with cutoff frequency near 5 mc (megacycles) are also presented.

D-C Characteristics

THE dependence[1,3] of the drain current on the drain and gate voltages is shown in Fig. 1(C). When a positive voltage is applied to the drain, with the source grounded, a current will flow from the source to the drain. If a negative voltage is applied to the gate, the drain current will decrease due to conductivity modulation.

When the magnitude of the gate voltage is increased to the pinch-off value W_0, the electron current will cease to flow, and the drain current is essentially the leakage current of the *pn* junction. On the other hand, if the gate voltage is kept constant, the drain current will increase with the drain voltage until a certain drain voltage is reached. After that, for further increase in drain voltage, the current remains relatively constant, indicating a high output impedance. It is to be noted that the gate voltage should always be kept negative; otherwise, excessive currents will flow, causing damage to the transistor. This is similar to the positive grid current of a vacuum tube.

Equivalent Circuit

Since the fundamental transistor action is a modulation of the drain current responding to a change in the gate voltage, the active part of the transistor can be best described by a transadmittance. This suggests an a-c equivalent circuit as shown in Fig. 2. It consists of the parallel combination of a passive, reciprocal network with elements y_{11}, y_{12}, y_{22} and an active current generator $(y_{21}-y_{12})v_G$. The relationship between drain and gate voltages and currents is:

$$i_G=y_{11}v_G+y_{12}v_D$$

$$i_D=y_{12}v_G+y_{22}v_D+(y_{21}-y_{12})v_G$$

The y parameters are complex, and change with frequency. They are measured by using a Tektronix *535* dual-beam scope, with the d-c bias kept at a drain voltage of 60 volts and a drain current of 5 milliamperes. (Because of high impedance levels, the y parameters cannot be measured accurately by a *GR* radio-frequency bridge. A detailed measuring technique is described in Appendix I.)

The results of the measurement of the conductance matrix elements are shown in Fig. 3. We can see that g_{11} decreases with frequency. Since g_{21} is essentially the transconductance of the active generator, it shows the frequency response of the transistor. g_{12} is negative and is smaller in magnitude than g_{11} and g_{22}. This ensures that the network consisting of y_{11}, y_{12}, y_{22} is a passive network. The measured results of the susceptance matrix elements are shown in Fig. 4. The fact that b_{12} is negative, and that the magnitude of b_{12} is less than b_{11} and b_{22} indicates that the susceptances are capacitive. When these matrix elements are combined, the passive network formed by y_{11}, y_{12}, and y_{22} can be reduced to an R/C (resistance-capacitance) network as shown in Fig. 5. The interelectrode capacitances C_{gs}, C_{gd}, C_{ds}, remain practically constant over the entire frequency range. For the transistor studied, C_{gs} is about 23 $\mu\mu f$ (micromicrofarads), C_{gd} is only 3 $\mu\mu f$, and C_{sd} is about 33 $\mu\mu f$. However, the resistances decrease with frequency. These decreases are probably due to the distributed nature of the device. The variation of the transadmittance y_m is also plotted in Fig. 5. It may be seen that the 3-decibel point, at which g_m equal b_m, occurs around 4.5 mc. This does not represent the highest frequency characteristics obtainable for a field effect transistor.[4] On the other hand, it represents the characteristic of a typical transistor fabricated, which has a reasonable cost and yield. If the physical dimensions of the transistor are reduced, the frequency response of the transistor can be improved at the expense of reduction in power rating and increase in cost. The maximum frequency of oscillation for a typical transistor described is around 10 to 12 mc. The input resistance remains relatively high, up to 500 kc. Beyond that frequency, the resistances decrease rapidly from several hundred thousand ohms to several thousand ohms. The results indicate that within a certain bandwidth the elements can be approximated by a constant R/C network, with the resistance taking a value at the center of the frequency band. This approximation results in great simplicity in the amplifier, oscillator, or other circuit designs.

Colpitts Oscillator

To illustrate the use of the equivalent circuit, a class-*A* Colpitts oscillator is designed. The objective is to design the tank circuit according to a specified frequency of oscillation and to predict the voltage swing and power output from the value of the load resistance. A schematic diagram of the oscillator is shown in Fig. 6.

By making use of the equivalent circuit, the loop equations for the circuit may be written. The determinant of the simultaneous equation is set to zero, some terms are neglected, and the conditions of oscillation are obtained; see Appendix II.

$$f\doteq\frac{1}{2\pi}\sqrt{\frac{C_d+C_g}{C_dC_GL}}$$

$$g_{11}\left(\frac{C_d}{C_g}\right)^2-(g_{21}+g_{12})\frac{C_d}{C_g}+\left(\frac{1}{R}+g_{22}\right)=0$$

Paper 56-252, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing January 23, 1956.

CHAANG HUANG and MELVIN MARSHALL are with Sylvania Electric Products, Inc., Woburn, Mass., and B. H. WHITE is with CBS Hytron, Lowell, Mass.

The authors wish to thank D. Unger and N. J. Golden, who are responsible for the development of the field effect transistors reported in this paper.

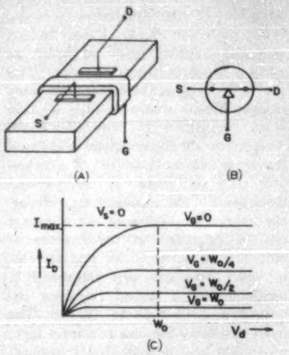


Fig.1 A—Field effect transistor. B—symbol. C—static characteristics

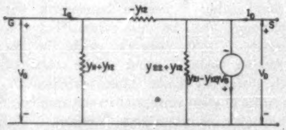


Fig. 2. Equivalent circuit of field effect transistor

From these two conditions, if one of the three tank circuit elements is chosen, the other two elements can be calculated. To check the accuracy of these conditions, experimental results have been compared to the calculated results. It is found that for $f=2$ mc and $L=140$ millihenrys, the calculated value of C_d/C_g is 0.63, and C_d equals 74 $\mu\mu f$. On the other hand, when the oscillator is tuned to 2 mc, the measured value of C_d/C_g is 0.618, while the value of C_d is found to be 60 $\mu\mu f$.

If it is recalled that the equivalent circuit is essentially resistive up to 2 mc, it is plausible that the power output and efficiency of an oscillator with an oscillating frequency less than 2 mc, can be directly predicted from the static characteristic curve as shown in Fig. 7. With the load resistance given, a load line may be drawn on the static characteristic curve. The load line intersects with the curves $V_G=0$, and $V_G=-33$ volts (the pinch-off voltage) at two points, giving two voltage swings. The smaller of the two swings is the drain voltage swing, v_D. From the voltage swing and load resistance the power output can be calculated. Since the d-c biases are known, the efficiency can also be calculated.

To verify the validity of these calculations, the power output and efficiency of a field effect transistor, the Colpitts oscillator is measured under the following conditions.

First, at a fixed bias point, $V_D=60$ volts; $I_d=6$ milliamperes, with the frequency of oscillation tuned to 2 mc; the load resistance has been varied from 2.2 K to 12 K. The measured and calculated results are found to be in good agreement, as shown in Fig. 8. This shows that for frequencies up to 2 mc, the amplitude of oscillation can be accurately predicted from the static characteristics. Furthermore, corresponding to each bias point, there is an optimum load to give maximum amplitude of oscillation. For the operating point of $V_D=60$ volts, $I_d=6$ milliamperes, $V_G=-12$ volts, the optimum load is about 6.2 K.

Second, the a-c power output and the efficiency of the oscillator are measured when the frequency is varied from 1 to 8 mc; see Fig. 9. With the operating point kept at $V_G=-12$ volts, $V_D=60$ volts and with a 6.2 K load resistance, the frequency of oscillation is tuned from 1 to 8 mc. It may be seen that when the frequency is higher than 2 mc, there is a decrease both in power output and efficiency. In other words, at frequencies higher than 2 mc, the power output and efficiency cannot be predicted directly from the static curves. The maximum frequency of oscillation for this circuit is found to be 11.5 mc.

Third, by maintaining a constant drain bias of 50 volts and a constant frequency of oscillation of 2 mc, the power output and efficiency are measured versus gate bias. (At each gate bias the load resistance is adjusted to the optimum value.) The measured results are compared with the calculated values in Fig. 10. As can be readily explained from the static characteristic curves, there is an optimum gate bias which gives maximum power output. This optimum bias occurs at $V_G=-12$ volts.

Fourth, when the transistor is operating at a constant gate bias, with the tank circuit tuned to a fixed frequency, an increase in drain voltage will shift the operating point to allow a larger amplitude of oscillation. Fig. 11 shows the measured and calculated power output and efficiency versus drain voltage with $V_G=-12$ volts, $f=2$ mc, and optimum load resistance.

From this study of the class-A oscillator, it is concluded that the tank circuit can be designed by the conditions of oscillation derived from the equivalent circuit. The power output and efficiency can be calculated from the static charac-

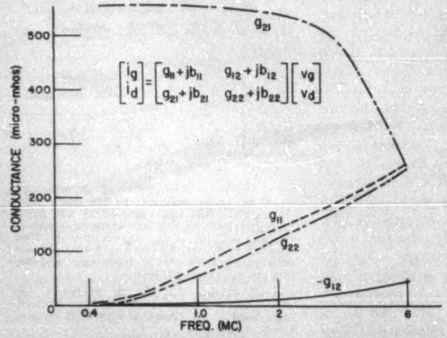


Fig. 3. Conductance matrix elements of a field effect transistor versus frequency

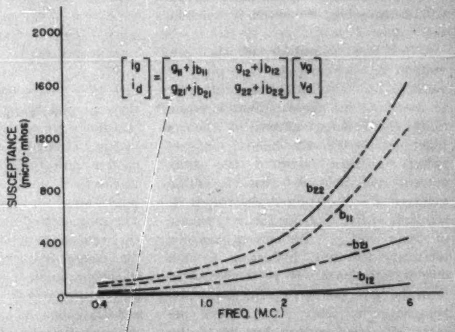


Fig. 4. Susceptance matrix elements of a field effect transistor versus frequency

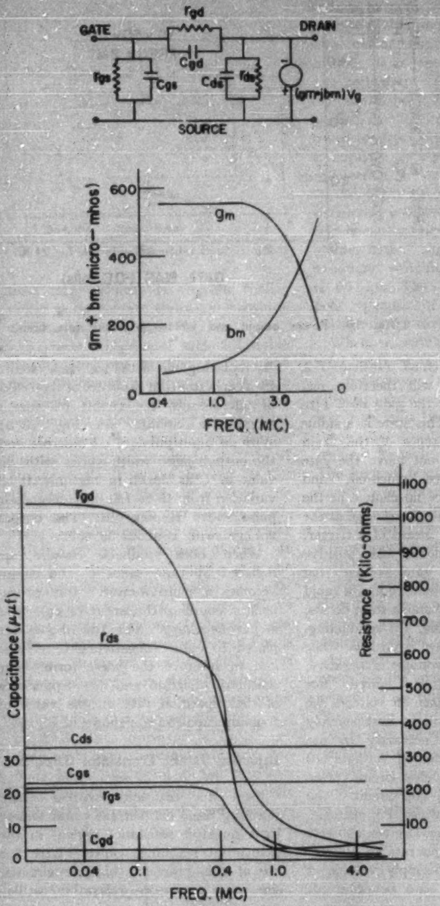
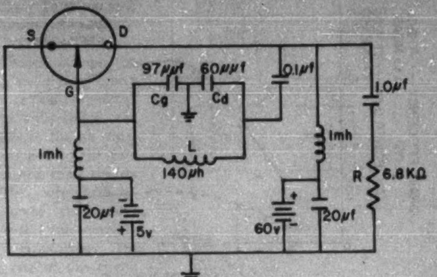


Fig. 5. Equivalent circuit elements of a field effect transistor



CONDITIONS for OSCILLATION

$$\omega = \sqrt{\frac{C_d + C_g}{C_d\, C_g\, L}}$$

$$g_{11}\left(\frac{C_d}{C_g}\right)^2 - (g_{21}+g_{12})\,\frac{C_d}{C_g} + \left(\frac{1}{R}+g_{22}\right) = 0$$

Fig. 6. Colpitts oscillator

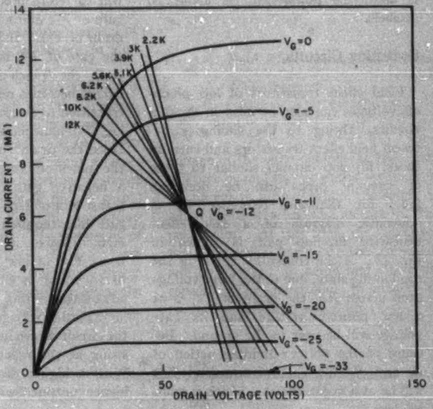


Fig. 7. Static characteristic

teristic if the frequency of oscillation is lower than 2 mc. This frequency is approximately one-half of the cutoff frequency at which $g_m = b_m$.

Power Rating

Although the highest power output data shown here is only 100 mw (milliwatts), output of more than 300 mw has been observed. The heat dissipation rating of the field effect transistors is 500 mw. If a class-*B* or *C* oscillator is designed, larger power output at 2 mc is possible. Furthermore, if the heat sink is improved, the same transistor could be made to dissipate 2 watts. This would make the field effect transistor unique in giving large output power at the range of 2 mc.

Neutralization

As shown in the equivalent circuit, the coupling between the gate and the drain is a simple parallel R/C combination. In a narrow frequency band, the inherent feedback caused by this R/C combination can easily be neutralized and unilateralized amplifiers thus achieved. (See, for example, reference 6.)

Application Utilizing High Input Impedance

The high input impedance of the field effect transistor makes it ideal in some applications, such as a preamplifier connecting to a high-impedance micro-

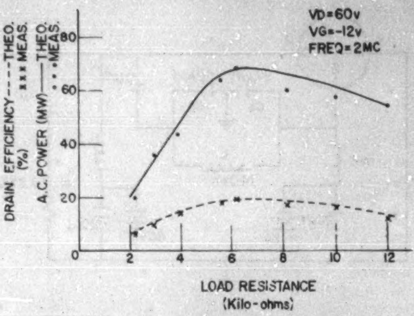


Fig. 8. Power output and efficiency versus load resistance

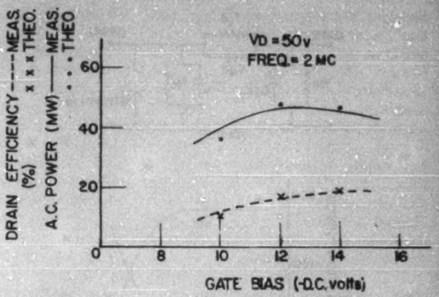


Fig. 10. Power output and efficiency versus gate bias

phone. However, in order to obtain reasonable efficiency, low pinch-off transistors are preferred (W_0 in the order of 10 volts). The typical input resistance ranges from 50*K* to 200*K*, the transconductance is about 500 micromhos. Thus, if a load resistance of 50*K* to 100*K* is used, the power gain is about 20 decibels.

Switching Circuits

Field effect transistors of low pinch-off voltages can also be used in flip-flop circuits. Owing to the similarity between field effect transistors and vacuum tubes, flip-flop circuits similar to those of vacuum tubes can be designed. (See, for example, reference 7.) A schematic diagram of a field effect transistor flip-flop pair is shown in Fig. 12. If the transistor *T*-1 is in the conducting state, due to the large voltage drop across R_{d1}, the drain of *T*-1 is at a low positive voltage and its gate voltage will be near to the ground. Because of the voltage-dividing action of R_1 and R_2, the gate of transistor *T*-2 will be at a negative value which is more than sufficient to pinch off transistor *T*-2. The drain of *T*-2 will therefore rest close to E_d, causing the gate of *T*-1 to rest near ground. This state is a stable one since small changes of the drain voltage of *T*-1 will not move the gate voltage of *T*-2 above "pinch-off," and will therefore produce no change at the gate of *T*-1 and no further change at the drain of *T*-1. If by means of a trigger, the gate of *T*-2 is moved more positive than the pinch-off gate voltage, the circuit becomes regenerative and a rapid transition to a second stable state occurs, i.e., *T*-1 nonconducting, *T*-2 conducting. Since the drain voltage is positive while the pinch-off gate voltage is negative, a negative gate bias is required. For transistors of high pinch-off voltage, the gate bias required will be unreasonably high. Hence, it is necessary to use transistors of low pinch-off voltages (10 to 15 volts) in a flip-flop circuit. One advantage of this flip-flop circuit is the large output voltage swing. For example, the circuit shown in Fig. 12 gives a 50-volt swing at the repetition rate of 100 kc. With a higher drain supply voltage, a higher output swing can be obtained. By replacing one of the feedback circuits by a capacitor *C* in the order of thousands of microfarads, the circuit becomes a monostable circuit; see Fig. 13(B). With an input pulse of 0.6 microsecond, the output pulse width varies with the value of *C* as shown in Fig. 13(A). A variation from 10 to 120 microseconds in pulse width is observed. The output voltage swing is about 50 volts.

When both feedback circuits are replaced by capacitance *C*, the circuit becomes a multivibrator. Due to capacitor coupling the negative gate bias is not necessary. The two diodes are placed to ensure negative gate voltage and to improve the wave form. The schematic diagram and the dependence of the repetition rate on the value of coupling capacitor are shown in Fig. 14.

Injecting Drain Transistor Circuits

Field effect transistors with a carrier injecting drain contact are characterized by a negative resistance region[3] in the gate input resistance curve. With the use of this property, bistable circuits, sine wave oscillators, relaxation oscilla-

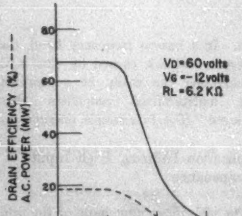


Fig. 9 (left). Power output and efficiency versus frequency

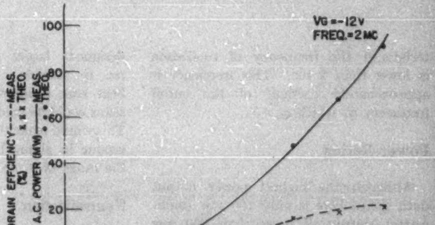


Fig. 11 (right). Power output and efficiency versus drain bias

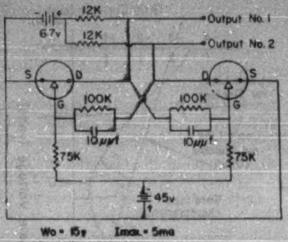


Fig. 12. Field effect transistor bistable circuit

tors, and monostable circuits similar to the point contact transistor switching circuits may be designed.[8,9]

A circuit diagram of such a bistable circuit is shown in Fig. 15(A). The drain is biased positive. If the gate terminals *GG'* are taken as the input terminals, a point-by-point measurement of the gate voltage and current characteristic can be obtained. A typical characteristic is shown in Fig. 15(C). (It has been pointed out that in the positive gate voltage and current quadrant, the characteristic would be similar to that of the double-base diodes.) It has three regions of operation: one positive low-resistance region, one negative resistance region, and one positive high-resistance region. It becomes apparent that if a proper gate load line is chosen, as shown in Fig. 15(C), three points of operation will be obtained. Among these points, *A* and *B* are stable points while *C* is an unstable point. At point *A*, the transistor is in the ON state with both the drain and gate currents large. At point *B* the transistor is in the OFF stage with both the drain and gate currents small and the gate voltage close to $-E_{G1}$. The OFF gate voltage will be dependent on the transistor characteristic. To modify this, a diode *D* is used which clamps the gate voltage to $-E_{G1}$, and the OFF point is raised to point *B*. A positive voltage applied to the gate will switch the transistor from the OFF to the ON state. A negative voltage applied to the gate will switch the transistor from the ON to the OFF state The source load resistance is included so that the bistable circuit can be triggered at the source to obtain better trigger sensitivity. The maximum switching rate of this circuit is 25 kc. The trigger sensitivity is poor because of the large pinch-off voltage of the experimental units.

When a parallel resonant circuit is used for the gate load, it presents a low d-c resistance. If the bias is selected properly, the d-c load line would intersect the gate input characteristic in the negative resistance region. Since this region is unstable, oscillation takes place. The highest frequency of oscillation is found to be about 300 kc. The output can be taken by inductive coupling to the tank circuit. The schematic diagram of a sine wave oscillator is shown in Fig. 15(B).

The source input characteristic is shown in Fig. 16. With this characteristic, a relaxation oscillator may be designed. First a load line is chosen which intersects with the input characteristic in the unstable negative resistance region. If a capacitor is placed across the source as in Fig. 16, relaxation oscillations will occur. The frequency of oscillation depends upon the R/C time constant. When the negative charge built up on the capacitor makes the source voltage more negative than the voltage of the valley point, the transistor will turn on and stay on until the capacitor discharges to a point where the source is more positive than the voltage of the peak point. At this time, the transistor will turn off. Once off, the capacitor will be charged up again and the cycle repeats. The maximum frequency of oscillation obtained was 6 kc.

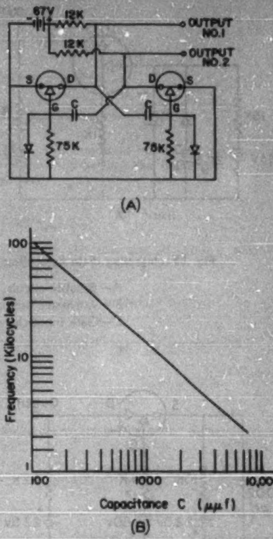
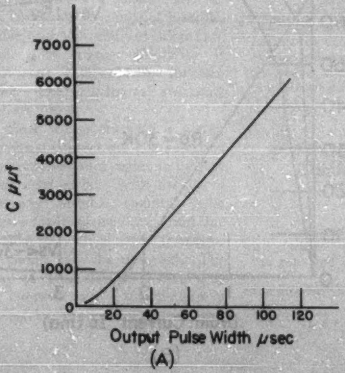


Fig. 14. Field effect transistor multivibrator

A—Multivibrator
B—Frequency versus capacitance



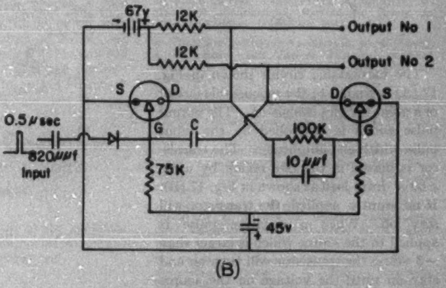


Fig. 13 Field effect monostable circuit

A—Pulse width versus capacitance
B—Monostable circuit

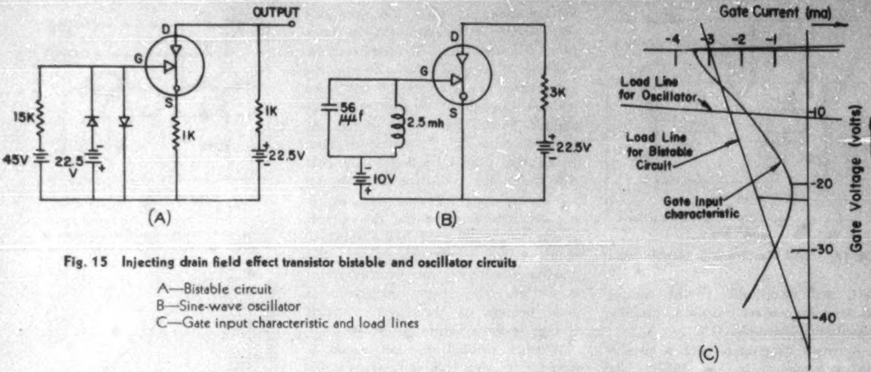

OUTPUT
D
G
S
15K
45V
22.5 V
1K
1K
22.5V
(A)
56 μμf
2.5 mh
10V
3K
22.5V
(B)
Gate Current (ma)
-4
-3
-2
-1
Load Line for Oscillator
Load Line for Bistable Circuit
Gate Input characteristic
-10
-20
-30
-40
Gate Voltage (volts)
(C)

Fig. 15 Injecting drain field effect transistor bistable and oscillator circuits

A—Bistable circuit
B—Sine-wave oscillator
C—Gate input characteristic and load lines

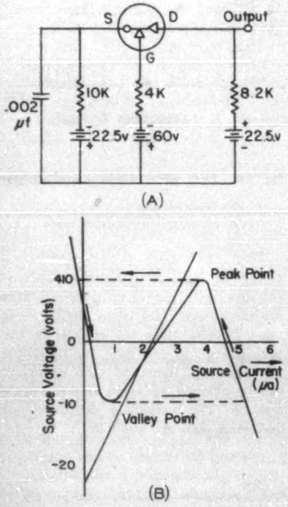

S
D
Output
G
.002 μf
10K
22.5v
4K
60v
8.2K
22.5v
(A)
410
Peak Point
0
1
2
3
4
5
6
Source Current (μa)
-10
Valley Point
-20
Source Voltage (volts)
(B)

Fig. 16 (left). Injecting drain relaxation oscillator

A—Relaxation oscillator
B—Source input characteristic and load line

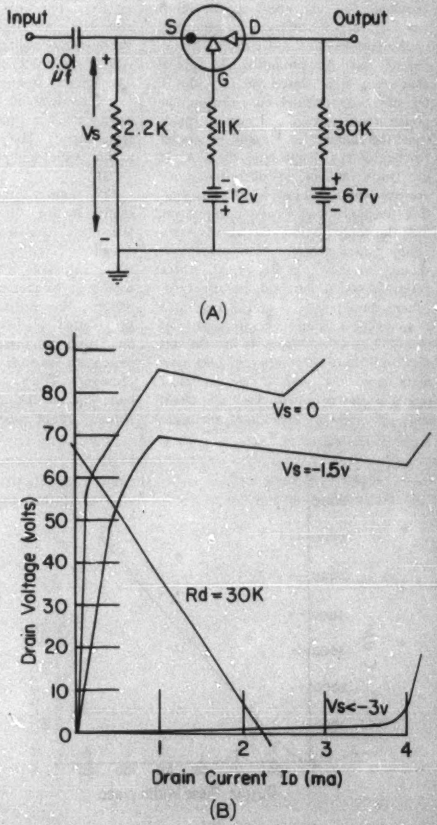

Input
0.01 μf
Vs
2.2K
S
D
G
11K
12v
Output
30K
67v
(A)
90
80
70
60
50
40
30
20
10
0
Drain Voltage (volts)
Vs = 0
Vs = -1.5v
Rd = 30K
Vs<-3v
1
2
3
4
Drain Current ID (ma)
(B)

Fig. 17 (right). A—Injecting drain field effect, transistor monostable circuit. B—Drain input characteristics and load line

The monostable circuit shown in Fig. 17 (A) is similar to the monostable circuit of a point contact transistor. The output pulse width is controlled by the input pulse width and amplitude. The transistor is biased in the OFF region by using a large drain load as shown in Fig. 17 (B). If no input is applied, the transistor will stay off. When a negative pulse is applied to the source which is larger than −3 volts, the transistor will turn on and stay on until the voltage on the source drops to a point where the gate diode is biased in the reverse direction, then the transistor will turn off and stay off until another pulse is applied.

Conclusions

An a-c equivalent circuit has been presented for field effect transistors on the admittance basis. It consists of the parallel combination of a passive R/C network and an active current generator. The values of capacitances remain relatively constant up to 5 mc while the resistances decrease with frequency. At frequencies less than 2 mc, the equivalent network is essentially resistive.

This equivalent circuit is used in the design of a Colpitts oscillator. The conditions of oscillation are derived by a circuit analysis, leading to the choice of the tank circuit. Experimental results are also presented on the oscillator performance under different operating conditions. It was verified that the a-c power output and efficiency of the oscillator can be calculated from the static characteristics even at frequencies up to 2 mc. The high dissipation rating of the field effect transistors make this type of transistor ideal for power oscillators in the frequency range around 2 mc.

Field effect transistors have high input impedances, especially for those with low pinch-off voltages. They are suitable for use in many applications, such as a preamplifier connecting to a high-impedance microphone, tuned or untuned intermediate-frequency or radio-frequency amplifier, cathode follower, and constant current device similar to a pentode. Since the feedback from the drain to the gate circuit is a simple parallel R/C network, neutralization can be realized easily.

Field effect transistors of low pinch-off voltage are also suitable for switching circuits. Since the electrodes can stand relatively high voltages (100 volts or higher), switching circuits can be designed to supply large voltage pulses. The design and circuit performance of bistable, monostable, and multivibrator circuits are presented. The trigger sensitivity is relatively poor compared to junction transistors. The maximum switching rate is about 100 kc.

Field effect transistors with a carrier injecting drain contact are characterized by a negative resistance region in its gate input resistance curve. With the use of this property bistable circuits, pulse amplifiers, relaxation oscillators, and sine wave oscillators with a single transistor in each circuit may be designed. The advantage of these circuits is their simplicity. However, due to the existence of minority carriers in this type of transistor the switching rate is much lower than that of point contact transistors.

In conclusion, the special features of field effect transistors are: 1. relatively high frequency of operation; 2. relatively high power rating, 3. high input and output impedance levels; and 4. high value of maximum d-c drain voltage. Because of these special features, field effect transistors are believed to be suitable for many applications in complement with junction transistors.

Appendix I. Measurement of Admittance Matrix Elements

Measurement of y_{11} and y_{22}

The signal generator is connected to the input terminal through a small current indicating resistance R. The output terminal is a-c grounded. Denoted by V_A and V_B the voltage developed across the two ends of the current indicating resistance is

$$y_{11} \text{ or } y_{22}=\frac{V_A\underline{/\phi_A}-V_B}{RV_B}-j\omega C$$

where ϕ_A is the phase angle of V_A with respect to V_B. The values of V_A, V_B and ϕ_A can be measured by using a Tektronix *535* dual-beam oscilloscope. C is the capacity of the scope lead across the input terminals. The value of R should be selected so that the real part of V_A is greater than twice the value of V_B. This eliminates the error introduced when subtracting two nearly equal quantities.

Measurement of y_{12} and y_{21}

The signal generator is connected directly to the input terminal. The output terminal is a-c grounded through a small current indicating resistance R. Denoted by v_C and v_D respectively, the voltages across the input and output terminals, are

$$y_{12} \text{ or } y_{21}=\frac{v_D-\underline{/\phi_D}\left(\frac{1}{R}+j\omega C\right)}{v_C}$$

where ϕ_D is the phase angle of v_D with respect to v_C. The values of v_C, v_D, ϕ_D can be measured by using a Tektronix *535* dual-beam oscilloscope. C is the capacity of the scope lead across the output terminals.

Appendix II. Condition of Oscillation for Colpitts Oscillator

The schematic diagram of the oscillator is shown in Fig. 6. The equation relating the gate and drain voltages are

$$(y_A+y_B)v_G-y_Av_D=0$$

$$[y_B+(y_{21}-y_{12})]v_G+y_Cv_D=0$$

where

$$y_A=-y_{12}-j\frac{1}{\omega L}$$

$$y_B=y_{11}+y_{12}+j\omega C_g$$

$$y_C=y_{22}+y_{12}+\frac{1}{R}+j\omega C_D$$

Thus, the condition of oscillation can be obtained by setting the determinant equal to zero

$$\begin{vmatrix} y_A+y_B & -y_A \\ y_B+(y_{21}-y_{12}) & y_C \end{vmatrix}=0$$

Equating the real and imaginary part of this equation to zero gives

$$\frac{g_{11}}{R}+g_{12}g_{11}-g_{21}g_{12}=(b_{11}b_{22}-b_{21}b_{12})+(C_gb_{22}+C_db_{11})-\frac{1}{\omega L}(b_{11}-b_{21}+b_{22}+b_{12})+\omega^2C_gC_d-\frac{C_g+C_d}{L}$$

and
$$g_{11}b_{22}+\frac{b_{11}}{R}+g_{22}b_{11}-g_{12}b_{21}-g_{21}b_{12}=\frac{1}{\omega L}\left(\frac{1}{R}+g_{22}+g_{12}+g_{11}+g_{21}\right)-\omega\left[g_{11}C_d+\left(\frac{1}{R}+g_{22}\right)C_g\right]$$

Numerical substitution for frequencies lower than 2 mc results in the following approximate equations

$$\omega^2=\frac{C_g+C_d}{LC_gC_d}$$

$$g_{11}\left(\frac{C_d}{C_g}\right)^2-(g_{21}-g_{12})\frac{C_d}{C_g}+\left(\frac{1}{R}+g_{22}\right)=0$$

References

1. A UNIPOLAR FIELD EFFECT TRANSISTOR, W. Shockley. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, Nov. 1952, pp. 1365–76.

2. UNIPOLAR FIELD EFFECT TRANSISTOR, G. C. Dacey, I. M. Ross. *Ibid.*, vol. 41, Aug. 1953, pp. 970–79.

3. JOINING SOLUTION AT THE PINCH-OFF POINT IN FIELD EFFECT TRANSISTORS, R. C. Prim, W. Shockley. *Transactions*, Professional Group on Electron Devices, *Ibid.*, Dec. 1953.

4. THE FIELD EFFECT TRANSISTOR, G. C. Dacey, I. M. Ross. *Bell System Technical Journal*, New York, N. Y., vol. 34, no. 6, Nov. 1955, pp. 1149–89.

5. A HIGH IMPEDANCE SILICON FIELD EFFECT TRANSISTOR, G. L. Pearson. *Bulletin*, American Physical Society, New York, N. Y., vol. 28, 1953, p. 7.

6. ELECTRONIC CIRCUITS AND TUBES (book), Cruft Laboratory Staff. McGraw-Hill Book Company, Inc., New York, N. Y., 1947, pp. 450–51.

7. ELECTRONICS EXPERIMENTAL TECHNIQUES (book), W. C. Elmore, M. Sands. McGraw-Hill Book Company, Inc., New York, N. Y., 1949, pp. 78–123.

8. TRANSISTOR TRIGGER CIRCUITS, A. W. Lo. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, Nov. 1952, pp. 1531–40.

9. TRANSISTORS IN SWITCHING CIRCUITS, A. E. Anderson. *Ibid.*, pp. 1541–58.

The Wisconsin Integrally Synchronized Computer—A University Research Project

J. L. ASMUTH
ASSOCIATE MEMBER AIEE

C. H. DAVIDSON
NONMEMBER AIEE

J. B. MILLER
ASSOCIATE MEMBER AIEE

D. S. NOBLE
NONMEMBER AIEE

A.K. SCIDMORE
NONMEMBER AIEE

IN THE fall of 1950, the University of Wisconsin began research in the design and construction of an automatic electronic digital computer to add to its servomechanism, analogue, and digital computer program. The project was undertaken for two reasons: 1. to obtain an additional research tool to supplement existing computing facilities such as the International Business Machines equipment operated by the university's numerical analysis laboratory and the electrical engineering department's high-speed electronic differential analyzer; and 2. to provide training for both students and staff in the problems of computer design, construction, maintenance, and use.

The essential features of the design were advanced by Amdahl.[1] As a result of an exposure to some of the early computers then operating at the Aberdeen Proving Ground, he became concerned over the bottleneck existing in the problem preparation phase of computing, and sought to incorporate in his design all possible features to simplify the operation of his machine from the user's point of view. His proposals were taken up by the department of electrical engineering and in the fall of 1950 the Wisconsin Integrally Synchronized Computer (WISC) was begun. A year and a half were spent in further study and design. No deadline for completion was set. The objective was to conduct a modest program of education, research, and construction, and to make the most of the opportunities for training young electrical engineers and acquainting them with some of the basic problems in this rapidly expanding area. Actual construction began in the spring of 1952, and now in the fall of 1955 testing and computing are alternately being carried on, and the WISC, Fig. 1, is about to begin its useful life. It is as far as known the only noncommercial large-scale computer built entirely without the support of the Federal Government.

Functional Design of the Computer

The computer itself is a fully automatic stored-program device of what is called intermediate speed. An over-all functional diagram is shown in Fig. 2. Like almost all such devices its information storage mechanisms are purely binary in character, and its internal arithmetic is therefore carried out entirely in the binary mode. Communicaton with the machine is by conventional decimal numbers, and conversion between binary and decimal is performed by stored subroutine.

In the early design stages considerable attention was given to the problem of speed versus complexity, and the apparent conflict between the goals of low cost and reliable high-speed operation. In particular it was recognized by Mr. Amdahl[1] that optimum performance would be obtained only when the speed capabilities of memory and arithmetic units were properly matched. Either performing parallel arithmetic operations on numbers stored on a relatively slow access drum, or combining serial arithmetic with a high-speed cathode-ray tube storage, would be an inefficient balance of cost of equipment versus speed of operation. Since over-all cost was an important consideration in this project, magnetic drum memory and serial arithmetic were chosen, but these were then carefully examined to determine the most efficient integration of their separate capabilities.

FLOATING POINT

As soon as operating experience with automatic computers began to be logged, it became apparent that the problems of fixed-point operation, which were relatively minor with hand-operated desk calculators, were tremendously magnified in automatic operation. So time-consuming did these scale factor adjustments become that a large number of computer users found it simpler and, in the long run, quicker to program the necessary steps to have the machine carry out all arithmetic operations in floating point, with automatic adjustment of the decimal point at each step, and to slow down the machine operating speed in favor of increased problem preparation speed. Such experience led to the decision to design into the WISC the additional circuits necessary to perform floating-point arithmetic automatically.

To the writers' knowledge, the WISC was the first electronic computer to propose such operation, and it is a decision which has never been regretted. The additional circuitry occupies perhaps 15 per cent of the computer, but the estimates of time saved by simplified programming range from 20 to 50 per cent of total problem time.

3-ADDRESS LOGIC

A 3-address logic was chosen for the WISC, since the word length needs to be great enough to include an exponent and its sign as well as the significant bits of a number with its sign. Whereas among experienced programmers the relative advantages of 1-, 2-, 3-, and n-address instructions will probably continue to be debated for many years to come, experience in the Wisconsin Numerical Analysis Laboratory strongly supports the contention that 3-address instructions are simpler to teach to beginning operators. As many people as possible should be able to use the WISC.

WORD STRUCTURE

With three addresses per order and floating-point notation in a number, the make-up of these two types of words presents a very good balance, as shown in Fig. 3. A number is represented as $\pm p \times 2^{\pm q}$, where p consists of 40 bits, q of 8, and 2 more are taken for signs. This converts at input and output into $\pm A \times 10^{\pm B}$ which is a 10-digit decimal number with an exponent range of $10^{\pm 75}$.

Paper 55-742, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 2-7, 1955. Manuscript submitted April 4, 1955; made available for printing November 7, 1955.

J. L. ASMUTH, C. H. DAVIDSON, J. B. MILLER, D. S. NOBLE, and A. K. SCIDMORE are with the University of Wisconsin, Madison, Wis.

The authors wish to express their appreciation to the entire staff of the Electrical Engineering Department, and particularly to its chairman, H. A. Peterson, without whose vision and understanding support the project could never have been carried through, nor indeed have come into being.

Fig. 1. The WISC

An order contains three 12-bit addresses, 4 bits to specify one of 16 fundamental operations, and 10 additional bits of which 6 are used only in an "extract" order. In either case, an additional 5 bits are allowed for switching time between words, making the total word length 55 bits. In the arithmetic operations the significant bits are separated from the exponent and signs, making a number composed of 40 bits plus 4 blank spaces. The lengths of these numbers are related by the cyclic memory requirement that forming 40 partial products of a 44-bit number takes the same number of digit times as storing 32 words each 55 bits long, 1,760 digit times.

Also shown in Fig. 3 is the order code for the WISC with an interpretation of the three addresses in each case. There are six arithmetic orders, eight orders which control the flow of computation, and one extract or logical multiplication order which allows the completely general modification of the contents of any register.

INTEGRAL SYNCHRONIZATION

In most fixed-point arithmetic, multiplication takes much longer than addition or subtraction, and division takes still longer. Consequently, it is common practice to sequence such a computer asynchronously, i.e., the successful completion of any one stage of the computation initiates the next. However, when the exponents have to be examined, the numbers shifted before and after adding, and the new exponent determined and formed, addition and subtraction times turn out to be more nearly equal to the times required for multiplication and division, and the gain in speed by asynchronous operation becomes much less. In fact, more time can be saved by making the maximum time necessary for any operation, arithmetic or otherwise, exactly the same, and by synchronizing all such operations with the memory cycle itself. In the WISC this works out in the following way.

As listed in the first column of Table I, a typical 3-address arithmetic order to the WISC requires that the following steps be performed:

1. The order must be located and read from the main storage. On the average this requires 1/2 memory cycle; to guarantee completion, 1 full cycle.

2. The addresses of the two operands *A* and *B* called for in the order must be located and their contents read. Since this is done simultaneously on two different readers, this requires an average of a 3/4 cycle, a maximum of 1.

3. The arithmetic operation on *A* and *B* called for must be carried out. For multiplication and division this has been adjusted to take 1 cycle. For addition and subtraction the time may be less but, to allow for the maximum amount of shifting in the worst case, nearly 1 full cycle is needed. One full cycle is allowed in all cases.

4. The result must be stored in the *C* address specified by the order. Time requirements are the same as in step 1.

To carry out any one order therefore takes 4 drum cycles on the WISC. However, since it can be guaranteed that each operation will be completed in the same time, all four operations can be carried out at once on four different orders. This procedure, called integral synchronization, results in the fourfold gain in speed shown in the remainder of Table I, and is the origin of the name Wisconsin Integrally Synchronized Computer.

One difficulty with the scheme shown in Table I was realized immediately, namely, the apparent impossibility of calling for the immediately preceding (or second preceding) result as an operand. However, a closer look at the timing shows that even this is possible by use of a special notation called short-memory coding.

The result of order n is guaranteed to be available from arithmetic by the end of cycle $n+2$. Although in the normal course an operand called for by order $n+1$ is located during cycle $n+2$, it is

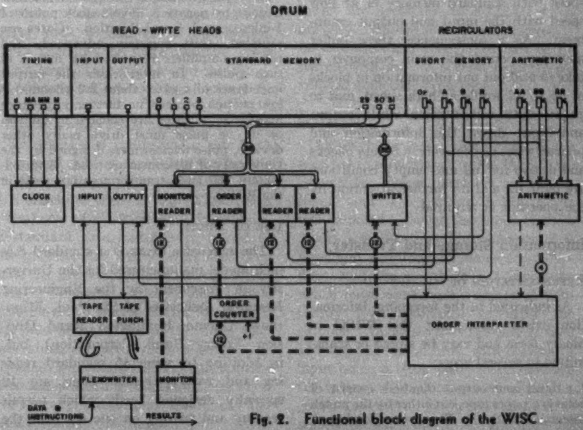


Fig. 2. Functional block diagram of the WISC

Table I. Integral Synchronization

Cycle No.	Order No.			
	n	$n+1$	$n+2$	$n+3$
n	Read order n			
$n+1$	Locate operands	read order $n+1$		
$n+2$	Perform arithmetic	locate operands	read order $n+2$	
$n+3$	Deliver result	perform arithmetic	locate operands	read order $n+3$
$n+4$		deliver result	perform arithmetic	locate operands
$n+5$			deliver result	perform arithmetic

not actually sent to arithmetic until the beginning of cycle $n+3$. Therefore, if the address of an operand in order $n+1$ is specified as short memory (binary 1000 0000 0000), the number emerging from the arithmetic recirculator *RR* at the beginning of cycle $n+3$ is returned immediately to arithmetic as operand *A* or *B*, as the case may be. Such special coding for preceding result has been used successfully in many computers, notably the card-programmed calculator.

If the result of order n is called for in order $n+2$, the machine finds itself trying to read and write in the same location during cycle $n+3$. This conflict is detected by the machine and a delay cycle is automatically inserted so that the result in question is first delivered, then located and used. Similar conflict detectors exist at other points in the machine's operation to eliminate the need for special attention to addresses by the coder; see section entitled "Reading Selector Circuits."

One additional logical feature deserves mention. Separate storage units of 32-word capacity each (subject to expansion) are reserved for input and output. These have additional circuitry associated so that they may communicate at high speed with standard memory or at slow speed with the input and output equipment. As a consequence, these buffer memories permit the main computer to take in and put out information in blocks of up to 32 words at high speed, and to go on with computation while output and input digest this information and prepare more respectively. Safety checks and delays for full and empty conditions are built in, and no further attention by the operator is required.

Information Storage and Transfer

REPRESENTATION OF "1" AND "0"

As indicated in the foregoing, information exists throughout the WISC in binary form and may be stored or transmitted in several ways:

1. Input and output numbers consist of holes in a paper tape, cut either by the punch associated with the Flexowriter keyboard, or by the punch attached to the output circuits themselves.

2. In the standard memory, information exists as magnetized cells on the surface of the drum, polarized one direction for "1," the opposite direction for "0." Not the least of the advantages of this type of storage is its nonvolatility with respect to power shutdown.

3. In electronic circuits information may exist or be transmitted in either of two forms: as one of two d-c levels, or as the presence or absence of a negative pulse at an expected time.

a. D-c information is obtained from the output of flip-flops used as storage, as counters, or as shift registers, or in special cases from manually set relays. Because of the very common use of a flip-flop output as a control voltage, the d-c levels at the plate of a flip-flop, +130 volts and +55 volts, are taken as a standard representation of "1" and "0," and manual inputs are converted to these levels for ease in combination. In this form, information can be used to control logical gates, to give visual information, or to control writing or punching operations. It has been found that in most cases the output of the triode flip-flop has a low enough impedance to permit direct coupling to the next stage; for control of more than one gate or for transmission to another chassis, the flip-flop is usually followed by a cathode follower.

b. Synchronization through the machine is done by negative 30-volt clock pulses of 1-microsecond (μsec) duration. Gates are sampled, shift registers advanced, and reading amplifiers reclocked by means of such pulses. In most cases the output impedance of a gate (about 2.2 kilohms) is low enough to need no further impedance matching; if gates are stacked in tandem or if one pulse must drive many other devices, pulse transformers designed by the University of Wisconsin are used. Approximately 260 such transformers are used in the WISC.

USE OF DRUM

The magnetic drum is a standard 8.5-inch model manufactured for the University of Wisconsin by the Engineering Research Associates of St. Paul, Minn. (now Univac Remington Rand Division, Sperry Rand Corporation), but, in addition to some 45 standard reading and writing heads, there are 12 specially designed heads which permit reading and writing on one track at the same time. Besides functioning as standard memory, the drum, as shown in Fig. 2, is used for input and output, for clock and timing tracks, for buffer storage between standard memory and arithmetic, and to provide a set of recirculating shift registers in the arithmetic operations themselves.

In a typical magnetic drum read-write head, both reading and writing windings are wound on one core, and either function can be carried on at any one time, but not both. In the special dual heads (Engineering Research Associates numbers *250A1* and *250B1*) a reading head is mounted 55 (in arithmetic 44) bits away from a writing head on the same track, and by means of elaborate shielding both operations can be carried on at one time, thereby allowing a single 55- (44-) bit word to be kept recirculating in this scratch paper or short memory.

These recirculators serve as buffers between standard memory and the arithmetic unit, between input and standard memory, and between standard memory and output, and perform the function of 55-bit static shift registers in each case. They are needed for such purposes as storing one operand until the second has been located, or for storing both until time for delivery to the arithmetic unit.

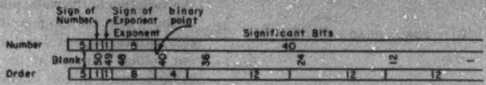
In the arithmetic unit itself, similar dual-head arrangements keep two operands and a result in constant recirculation with one additional modification. If the number being recirculated is read from the drum one bit early and passed through two stages of shift register, it can be tapped off from any of three different places to cause recirculation normally or with a shift right or left of one bit per recirculation cycle. Fig. 4 shows a dual head with its output fed back to its input through such a selector. This process can be extended as far as desired, which in the case of the WISC is two left shifts or four right shifts per cycle.

Structural Design of Computer

UNDERLYING PHILOSOPHY

The design of a complex piece of electronic equipment such as a digital computer admits a tremendous variety of possible detailed circuits for accomplishing specified objectives. A redeeming feature of this complexity, however, is the fact that even the most complicated circuit can be broken into large numbers of basic units or "building blocks" where the number of different types of such units can usually be made quite small. It was realized early in the WISC project that complex structure could only be

Vol.
Pt. 1
1956



Code	Symbol	Operation			Order Type	A		B		C	
0	I	Input	00		0000	00	Address for first word	00	Address for last word	00	Address of next order
1	E	Extract	00	00 x^6	0001	00	Address of operand A	z^2	y^6 | z^4	00	Address for result
2	M	Multiply	0n		0010	s0	"	s0	Address of operand B	r0	"
3	D	Divide	0n		0011	s0	"	s0	"	r0	"
4					0100						
5	TU	Unconditional transfer	0t		0101					00	Address of next order
6	H	Halt	0h		0110			00		00	"
7	O	Output	00		0111	00	Address of first word	00	Address of last word	00	"
8	A	Add	0n		1000	s0	Address of operand A	s0	Address of operand B	r0	Address for result
9	AA	Add absolute	0n		1001	s0	"	s0	"	r0	"
a	S	Subtract	0n		1010	s0	"	s0	"	r0	"
b	SA	Subtract absolute	0n		1011	s0	"	s0	"	r0	"
c	TZ	Transfer on zero	00		1100	s0	"	s0	"	00	Next order if A − B = 0
d	TZA	Transfer on zero absolute	00		1101	s0	"	s0	"	00	Next order if |A| − |B| = 0
e	TN	Transfer on negative	00		1110	s0	"	s0	"	00	Next order if A − B < 0
f	TNA	Transfer on negative absolute	00		1111	s0	"	s0	"	00	Next order if |A| − |B| < 0

Fig. 3. Structure of numbers and orders

h=0; Unconditional halt
h=1; Halt if switch is set
n=0; Normal arithmetic operation
n=1; Do not shift result
r=0; Normal operation
n=1; Do not deliver result to storage
s=0; Normal operation
s=1; Use preceding result for this operand
t=0; Unconditional transfer
t=1; Transfer if switch is set
x=First bit to extract
y=Place to insert first bit
z=Number of bits to extract
0=Plus
1=Minus

achieved by this procedure of assembling basic units, and that these units could be used as the basis for such building only after thorough design, analysis, and testing. Such a procedure reduces the number of uncertain elements in trying out any new circuit, and also greatly reduces the amount of "breadboarding" which is such a necessary part of electronic circuit design.

A second guiding principle was established in the matter of reliability. In spite of the desirability of small size and low number of tubes in any such computer, where the incompatability of these goals with the unquestioned need for reliability necessitated a choice, the choice was always made in favor of the more reliable circuit (where this could be determined).

Time constants must be carefully watched at all times, both to insure adequate recovery and to guard against the possibility of harmful oscillations when performing computations with unpredictable periodicities. In the same way, transformers and inductors must be critically damped or otherwise prevented from responding to any possible resonances.

One compromise with ultimate reliability (many are necessary of course in any realistic circuit) is the practice of using resistance-capacitance coupling to differentiate the output of one flip-flop into another. Although it is recognized that gating such a transfer of information would probably be more reliable, great circuit savings are gained, and the standard of operation achieved has justified the choice.

A compromise which has not been made is the practice of controlling a gate with a d-c level subject to change during the act of sampling. This can easily occur in a machine such as the WISC with several sets of clock pulses. Some early substandard gate pulses were traced to this practice and every effort has been made to prevent its occurrence, intentional or otherwise. Closely allied to this is the so-called suicide circuit in which a gate pulse shuts off one of the direct voltages which permitted its own generation. Although superficially this circuit appears to be self-protecting, unreliable operation of a sporadic nature may result, and the circuit has not been used.

FLIP-FLOP

The flip-flop used in the WISC is an essentially conventional Eccles-Jordan circuit, as seen in Fig. 5; the 1-kilohm grid resistor is added to suppress parasitic oscillations and is the only unusual feature. Since frequency requirements are eased by the relatively low 100-kc pulse rate, triodes with comparatively high grid circuit impedances can be used, effecting a significant economy in both number of tubes and power supply requirements. The circuit is a true trigger in the sense that the clock or gate pulse lasting less than 1 μsec initiates the transition, which is then maintained by the energy stored in the 50-micromicrofarad speed-up capacitors. The actual transition takes about 1.5 μsec; i.e., 1.5 μsec after the beginning of the initiating pulse, the plate of the side being cut off has achieved more than 95 per cent of its ultimate rise. To allow for loading effects and tube variations a minimum delay of 3.3 μsec before sampling the new information is maintained, or the state of the flip-flop is reversed.

In conventional or 2-input flip-flops, pulses are coupled in unilaterally by means of small capacitors to remove any inherent d-c bias. When flip-flops are coupled bilaterally, as in a counter, diode input coupling is used to prevent positive going transients from affecting the counter. Under these conditions the counter chains are sufficiently stable so that intermediate stages may be tapped to provide d-c input to gates or coincidence detectors without affecting the operation of either prior or subsequent stages.

All flip-flops and counters in the WISC use type-*6J6* tubes; where cathode followers are necessary, type-*12AU7* tubes are used. Should subsequent changeover be desired, these tube types can be replaced with any of several recently developed computer-type tubes.

GATES

Although d-c gates using crystal diodes are used to some extent in the WISC, the commonest logical operators are the 4-input "and" gates shown in Fig. 6. It will be noted however that three of these inputs are d-c enabling voltages of a level and character available from flip-flops, and the fourth is a 1-μsec pulse which must come from a source of sufficiently low impedance to drive the cathode. In most cases the gate is used as a 3-input logical unit, with the fourth input being a clock pulse to synchronize the information. The output of such a gate is a negatively going pulse of approximately the same magnitude and duration

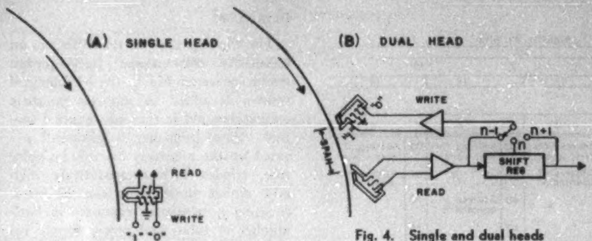


Fig. 4. Single and dual heads

as the clock pulse and, hence, is commonly used to set flip-flops and shift registers, and to drive counters. Upon occasion, however, when it is necessary to combine more than three inputs without additional delay, two gates may be stacked in tandem through impedance-matching pulse transformers. When so connected, the input pulse at the second cathode differs negligibly from a clock pulse.

As these gates are used, there is a similarity to the analogue or Kirchhoff adder in that one of four possible d-c levels may exist at the control grid of the gate. However the gate is always biased so as to pass a pulse only when all three inputs are high, for reasons which will be shown in the following.

Consider the three inputs A, B, and C in Fig. 6 as coming from the plates of three flip-flops. Each flip-flop may be at a potential of +130 volts (representing a "1") or +55 volts (representing a "0"). The control grid of the gate assumes the arithmetic mean of the three input levels and may, therefore, have a potential of +55, +80, +105, or +130 volts. If the cathode is normally biased at +150 volts but is driven down by negative 30-volt pulses, the cathode will be driven into the conduction region and the potential of the plate will drop accordingly, only if $ABC=1$. If only two of the three inputs are high, the grid will be at least 15 volts below the cathode at all times, which is well below cutoff. Gates with as many as five and even six such inputs were tested and found to operate quite well, but a maximum of three inputs was decided upon to reduce the need for critical control of pulse magnitudes, flip-flop loads, and the d-c gain of the cathode followers.

Now it is tempting to try to bias such a gate so that it will have the properties

$p(AB+BC+AC)=e$ (any two)

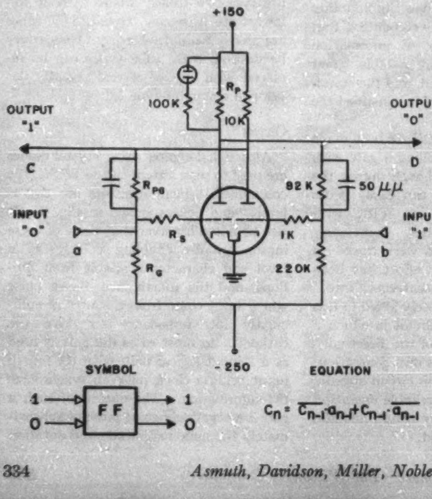
or even

$p(A+B+C)=e$ (any one)

This is dangerous, however, because of the variation in the static plate resistance of the side of the flip-flop under conduction with consequent variation in the output voltage achieved at the plate. *6J6* tubes as used in this circuit commonly range between 50 and 65 volts in this respect with occasional very good or very bad tubes giving values as low as 40 volts or as high as 70 volts. Consequently, if one or more of these conduction voltages is allowed to affect the grid control level which will permit a pulse to pass, variation with time and from tube to tube can seriously alter the discrimination of the tube and the quality of the output pulse.

Such a variation, however, does not occur in the plate voltage on the cutoff side, which is determined only by the potential divider effect of R_p, R_{pg}, and R_g. This voltage has been found to be constant at 132 volts with a fluctuation of less than 2 volts in the flip-flop circuit used, based upon approximately 100 measurements. If these voltages then are the only d-c levels which will enable the gate to pass a pulse, very high discrimination and very consistent output pulses may be obtained even with three inputs to a single gate. Where two gates are sampled with the same pulse, a *6J6* tube is used as a gate pair; if separate cathode pulsing is required, *12AU7* tubes are used.

D-c amplifiers, negators, and cathode followers of essentially conventional design also appear in the WISC, as do variations of a pulse amplifier which is used variously to supply clock pulses, to record on the drum, and to reshape other pulse information.



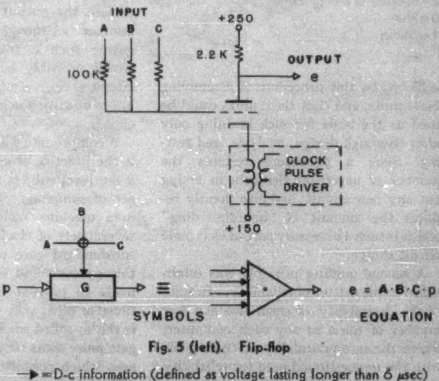


Fig. 5 (left). Flip-flop

→ = D-c information (defined as voltage lasting longer than 6 μsec)
▷ = Information in pulse form
These symbols also apply to Figs. 6, 7, and 9

Fig. 6 (above). Pulse gate

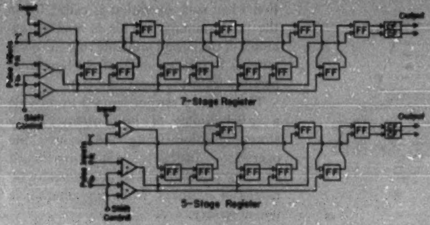


Fig. 7. Serial-parallel shift register

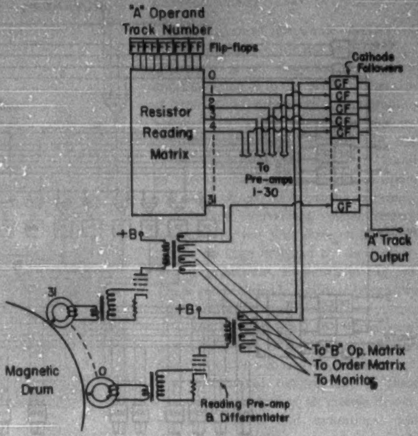


Fig. 8 (right). Reading selector circuits

TIMING

The use of a magnetic drum as the principal storage and the gearing of logical and arithmetic operations to the speed of writing digits on the drum greatly simplifies one problem in that it removes the necessity of trying to synchronize the drum to an externally controlled frequency. All timing signals for the WISC, including the clock pulses themselves, are generated from reference marks recorded directly on the drum. If the drum speed should change within reasonable limits, all repetition rates change with it, and no harm is done.

Timing signals are distributed about the WISC in two ways, i.e., as negative going 30-volt 1-μsec pulses, and as d-c enabling voltages for gate control, lasting 1-digit time (10 μsec) or less. Since the source of clock pulses applied to the cathode of the gate must have high power and low impedance, these are generated in a special clock section and supplied to all sections of the computer at +150-volt bias.

Signals to mark other intervals are frequently wanted, however, and the following set is commonly available throughout the machine:

	Intervals, Digit Times	Pulses per Revolution
MA, minor arithmetic....	44	40
MM, minor memory	55	32
M, major	1,760	1

Pulses at *MA*, *MM*, and *M* times are needed less commonly and, in many cases, can be generated on the chassis where wanted by applying a d-c enabling signal to an unused gate tube or, frequently, to an unused input of the gate where the timing signal is being combined with other control information. Since advantage is taken of the five blank spaces between words to perform many different operations, several sets of *M* enabling voltages are available, occurring at slightly different times within this 5-digit time interval.

Although the pulse repetition rate is described as approximately 100 kc, as determined by the drum revolution rate, the use of pulse gates makes it highly undesirable to wait until digit time $n+1$ to sample conditions set up by digit pulse n if this can be accomplished sooner. It was mentioned in the discussion of flip-flops that it can be accomplished in less than 10 μsec; therefore, the time interval of 1-digit time has been broken into three approximately equal sections, and three complete sets or families of clock pulses called α, β, and γ are provided, approximately 3.3 μsec apart. This constitutes what is known as a 3-phase timing system. Note that on any one clock-pulse line, however, pulses still occur only every 10 μsec.

The interval of 3.3 μsec was chosen not only because it allowed adequate time for flip-flop transitions with a reasonable safety factor but also because it corresponds very closely with the inherent delay in writing, reading, and shaping of drum signals, thereby permitting writing pulses to be coincident with one of the families of clock pulses.

SHIFT REGISTERS

The shift register is a versatile device much used in computing machines, whether its elements are flip-flops, magnetic cores, ferroelectric cells, or any other form of binary storage device. Its functions are many and varied, and include storage, shifting (as the name implies), delay, and serial-to-parallel and parallel-to-serial conversion.

Information may be shifted down the line of a register by sequentially pulsing each stage if an adequately elaborate sequencing system, enough input connections, and sufficient time are available. Instead it is common practice to pulse the several stages simultaneously, and to cause the entire word stored in the register to advance at the same time. Such a procedure imposes an extra memory requirement, however, for the information in each stage (except the last) must be stored somewhere while the following stage is being cleared. This storage may consist of small capacitors, short-delay lines, or auxiliary flip-flop stages constituting a second register. The last of these techniques is in fact frequently used, and in elaborate circuits may be gated to permit bidirectional shifting.

Two types of flip-flop shift registers are used in the WISC. Where it is desirable that the information be available at all intermediate states of the register, conventional dual registers with capacitor coupling between the stages rather than gates are used. This circuit has the disadvantage, however, of requiring four half-tubes per stage.

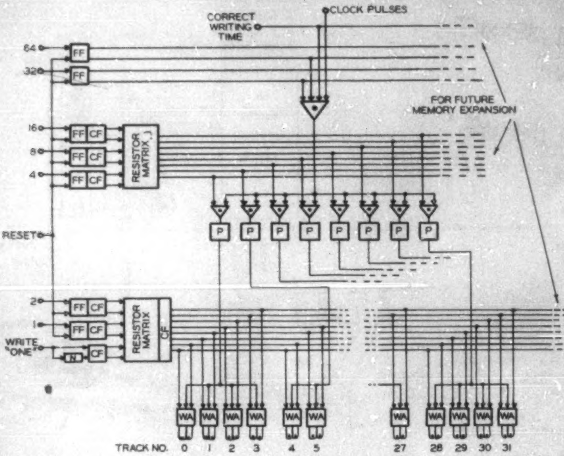


Fig. 9. Writing selector circuits

Where the output of the last stage is all that is desired, considerable saving in tubes can be effected by using the serial-parallel shift register, as shown in Fig. 7. Since the timing system in the WISC is 3 phase, as described in the foregoing, three different flip-flops can be pulsed sequentially in 1-digit time. Such a 3-stage register only shifts two different bits of information, however, since, upon completion of the third pulse (or any other), two stages contain the same information. As compared with conventional shift registers containing a complete set of temporary storage stages, a shift register of this type requires approximately 25 per cent fewer tubes.

Reading Selector Circuits

Fig. 8 shows the circuits used for reading out of the main storage section of the drum. The reading selection system must furnish four fully independent simultaneous outputs: one for reading the next order, two for locating the operands called for, and one for monitoring purposes. Although they are normally different words, any two or more of these outputs may be the same.

The selection of a desired word on one of the outputs is done in two steps: 1. the desired track on the drum is connected to that output; and 2. the correct word of the 32 is selected by its position in time measured from a mark pulse on the drum. Only that part of the circuit used for the track selection is shown in the figure.

The information written on the drum is amplified by a pentode preamplifier which, by using a transformer in the plate circuit, also accomplishes differentiation of the signal. These transformers have four secondary windings, one for each of the outputs. The d-c potentials on the return end of the individual secondaries are set by a selector matrix consisting of 160 resistors. Of a given set of 32 windings associated with an output channel, only one is at a level of approximately +130 volts. The other windings are from 15 to 75 volts below this level, with the result that only the signal which is biased at the high level appears at the output of the cathode follower.

This signal is then further amplified and, by means of a reclocking gate, is phase-detected and resynchronized. The desired word of those on a given track is selected by means of a coincidence detector which determines coincidence between the last five bits of the address of the word and a signal written on a control track of the drum which identifies each of the 32 angular positions of the drum.

The reading circuits must be capable of recovery from the very large signals caused by writing onto the drum within 40 μsec (4-digit times) to permit reading a word out of the next angular position. Simultaneous reading and writing on the same head (or even on adjacent heads) cannot, of course, be permitted, and detection circuits are built into the machine which allow an extra cycle of operation when such operation is called for in the programming. (This is another example of the design principle that the computer is to take as much of the burden as possible from the programmer.)

The monitor output is used to provide a visual display of the contents of any storage position independent of machine operation. In this case the desired track and time position are entered manually by means of switches, and programming is not necessary.

Writing Selector Circuits

The problem of the standard memory writing system is to select one of 32 tracks and, upon command, generate a writing pulse in either the "1" or the "0" side of the high-impedance writing coil in the head associated with that track. There are, therefore, seven pieces of binary information available as inputs: five to determine the track, one which selects the proper *MM* time, and one which controls whether the bit to be written is a "1" or a "0." The last of these must be capable of an arbitrary sequence of changes at 10 μsec intervals during writing.

The recording current pulses, supplied by *6AQ5* miniature beam power tubes, have a magnitude of 200 milliamperes for one μsec. Both "1"s and "0"s of a word are written but, because the track is saturated in the "0" direction, the readout signal for a "0" is considerably smaller than for a "1" and is, of course, of opposite phase.

Since one writing tube is connected to each half of the center-tapped write winding, there are 32 pairs of writing tubes required. The driving voltages for each pair of writing tubes are coupled from *6J6* pulse gates by means of 1-to-1 pulse transformers.

In the over-all block diagram of the standard memory writers and track selectors, Fig. 9, the pulse amplifiers *P* reshape and provide a low output impedance for the clock pulses which have already passed through two gates in tandem. The gate grid voltages are derived from resistor matrixes controlled by the binary track information in the flip-flops.

Coincidence Detectors

In the input-output section are six 5-bit parallel coincidence detector units, each consisting of a 5-stage binary counter and ten cathode-follower-driven diode coincidence detectors. These units are identical so that one spare unit serves for all. Three are used in input; three in output. In input, for example, one counter is used to indicate the *MM* position on the drum,

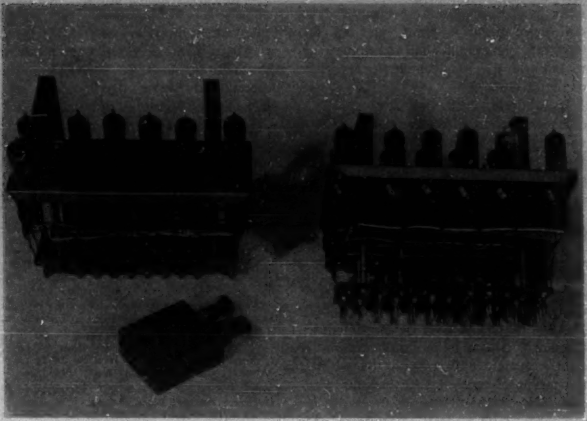


Fig. 10. Plug-in chassis

another to count the words read in from the tape to input memory, and the third to count the words delivered from input memory to standard memory.

When, and only when, all five pairs of bits of the two numbers being compared agree, coincidence is detected and a gate enabling voltage is made available through a cathode-follower driver to the appropriate gates, indicating that a word may be read in or out as the case may be.

The rise time of the gate enabling voltage from the coincidence detector is approximately 1 μsec which in most cases permits sampling one phase time (3.3 μsec) after the counter is advanced. However, in critically timed circuits, these gate enabling voltages are not sampled until two phase times (6.7 μsec) later to insure positive sampling.

Each coincidence detector unit employs 12 twin triodes: five in the binary counter, five as cathode-follower drivers to the coincidence diodes, one as a reset tube, and one as a combination negator and cathode-follower driver for supplying the coincidence gate voltage. In addition there are 20 type-*1N38* germanium diodes: ten for the coincidence detection, and ten for diode coupling in the counter circuit.

When the two numbers to be tested for agreement are in serial form, they may first be registered in a staticizer and the coincidence detector described in the foregoing may be applied, or a serial coincidence detector may be used. In such a device each of the two numbers is registered, bit by bit in time sequence, by a flip-flop. Each digit time these two flip-flops are examined for disagreement. If by the end of the interval of interest no disagreement has been detected, the two numbers must have been the same. Perhaps this device should be called an anticoincidence detector?

The detection of disagreement is accomplished by a pair of gates, one for each of the two ways in which a pair of binary digits can disagree. The two gates have a common output which is the input to a third flip-flop. At the beginning of the interval of interest this third flip-flop is reset, to be set only if there is a disagreement between a pair of corresponding digits.

This coincidence detector is capable of being tested for coincidence or lack of coincidence as soon as the last of the digits to be compared has been examined. The use of a parallel coincidence detector, in conjunction with a pair of shift registers to staticize the information, would not make coincidence information available any sooner, and would involve a substantially greater number of tubes, which would have to be increased with the number of digits in a number.

In addition to the four tubes constituting the serial coincidence detector, two timing signals are needed to mark the beginning and end of the interval of interest, one to sample the information in the flip-flop which records lack of coincidence, and one to reset this flip-flop. The number of tubes here is independent of the number of digits in a number.

PHYSICAL PLANT

As with most electronic devices of similar complexity, the use of small plug-in packages greatly simplifies construction and maintenance in the WISC. The size of these units represents a compromise between the opposing pressures of compactness and logical isolation. In the WISC, single and double chassis of 12 and 24 tubes respectively were decided upon, and these chassis have proved extremely flexible and convenient. Some experimental work has been done on a smaller 2-tube package designed by the University of Wisconsin, as shown in Fig. 10. These are intended to facilitate laboratory experiments accompanying course instruction, and to provide a 1-to-1 correspondence between circuit elements and a logical diagram in breadboard experimentation. The mechanical construction of these chassis (and indeed of all chassis, frameworks, cabinets, and mountings) was done by students or members of the university staff. There are 80 single and 14 double chassis in the computer circuits proper, in addition to 100 tubes in nonstandard chassis, such as power supplies, monitor facilities, etc. Altogether there are about 1,400 tubes in the WISC.

With the exception of the motor-generator set, the WISC is contained in four separate cabinets, each with its own heater supply and direct-voltage regulator units. The magnetic drum, the control circuits, reading and writing amplifiers, and the various clock pulse generators are located in the largest of these, which also houses master power switches and a running time meter. There are two small cabinets connected by cable to the large cabinet, one housing the arithmetic and the other the input-output circuits. The remaining cabinet is the operating console which contains the monitor, tape reader and punch, and associated controls.

Each cabinet has cooling blowers and is mounted on casters for flexibility in physical arrangement. The total floor area occupied excluding the motor-generator set is approximately 35 square feet.

Power is supplied to the machine from a 230-volt 3-phase line. The heater supply transformers in general have their secondaries isolated from one another so that heater-to-cathode voltages may be kept at a minimum in the various plug-in chassis by biasing the heater secondaries. Each cabinet has its own filament supply.

The direct current is supplied from a single motor-generator set which consists of one 3-phase driving motor and two d-c generators. One d-c generator supplies all of the positive-to-ground voltages and the other the negative-to-ground voltage. There are three positive voltages of +250, +150, and +130 volts respectively.

It has not been found necessary to regulate the +250-volt supply since it is used primarily in generating pulses. It is, however, supplied in isolated and filtered channels to the various logical units of the machine, the isolation and filtering being done as near as possible to the point of use. The +150-volt supply which is used as flip-flop plate supply, and which therefore furnishes critical control of gates, is electronically regulated with the use of conventional regulators so distributed that each supplies a small group of chassis. Eighteen such regulators are used in the complete machine, and give excellent isolation as well as regulation.

In regulating the negative 250-volt supply, conventional electronic regulators would prevent employing the same isolation scheme used for the +150 volts. The d-c generators are connected in a 3-wire system, and the common line is both direct current and signal ground. To achieve isolation, therefore, a series electronic regulator unit must have its regulator tube located in the minus line. The cathode of this tube must then be at a still more negative point, and special electronic bias supplies had to be provided to bias the d-c amplifier in the negative regulator circuits.

In addition, each section of the machine is supplied with a regulated +130-volt small-power supply line for additional gate bias purposes. The total power consumption of this machine is approximately 10 kw.

Results

In its nearly 5 years of operation, the digital computer project has paid off manyfold. A working computer of unique design has been developed, outstandingly simple to program, and sufficiently moderate in cost to be within the means of many institutions and organizations hitherto unable to justify a computer in its budget. Results in education and experience acquired have been invaluable. Formal training in both computers and computing has been provided in course instruction and graduate research. Some by-products of research have appeared as bonuses: a complete line of home-wound and potted pulse transformers, and several new types of plug-in package circuits for modular construction. Last but not least, the accumulated knowledge places the University of Wisconsin in a uniquely advantageous position for continued research in connection with computing machines, computer components, and computing in the years to come.

References

1. THE LOGICAL DESIGN OF AN INTERMEDIATE SPEED DIGITAL COMPUTER, **G. M. Amdahl.** *Ph.D. Thesis*, University of Wisconsin, Madison, Wis., 1951.

2. THE DESIGN OF AN ARITHMETIC UNIT FOR AN ELECTRONIC DIGITAL COMPUTER, **J. L. Asmuth.** *Ibid.*, 1955.

3. MATHEMATICAL PROCEDURES AND TESTING METHODS FOR AN INTERMEDIATE SPEED DIGITAL COMPUTER, **C. H. Davidson.** *Ibid.*, 1952.

4. AN ALGEBRAIC THEORY FOR USE IN DIGITAL COMPUTER DESIGN, **E. C. Nelson.** *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., EC-3, Sept. 1954.

5. THE DESIGN OF AN ORDER INTERPRETER FOR AN ELECTRONIC DIGITAL COMPUTER, **D. S. Noble.** *Ph. D. Thesis*, University of Wisconsin, Madison, Wis., 1954.

6. ARITHMETIC OPERATIONS IN DIGITAL COMPUTERS (book), **R. K. Richards.** D. Van Nostrand Company, Inc., New York, N. Y., 1955.

Transistor Analogue Computing Amplifiers for Flight Simulators

ROBERT C. WEYRICK
ASSOCIATE MEMBER AIEE

A WIDELY used flight-simulation method for flight trainers is the use of an analogue computing system employing a modulated 400-cycle-per-second (cps) carrier. A recent paper[1] described four basic computing components that have been developed for use in computing systems of this type. Two of these components are vacuum-tube amplifiers. These are: 1. a summing amplifier of the voltage feedback type having a 1-per-cent accuracy of summation; and 2. a power amplifier for servomotor operation. This paper describes the design and characteristics of two transistor amplifiers that are directly interchangeable with these vacuum-tube amplifiers.

Summing Amplifier

To provide for interchangeability and direct comparison, the performance requirements of the transistor summing amplifier are the same as those previously established for its vacuum-tube equivalent. These requirements are:

1. The amplifier should produce a balanced output of 25 volts rms across a load consisting of four 20,000-ohm potentiometers (two on each output) or a tuned resolver. This condition requires a power output of 125 milliwatts.

2. With the feedback resistor and each input resistor having a value of 500,000 ohms, the amplifier should sum a maximum of four inputs without introducing an error of more than 1 per cent. This condition requires a minimum effective gain of 40 decibels where effective gain is defined as the open-loop gain from one input to the output, and with the feedback resistor and the other three input resistors grounded through low-impedance sources.

3. The amplifier should remain stable when the shunt capacity of the resistive load is 0.001 microfarad or less for each potentiometer.

The schematic diagram of the transistor summing amplifier is shown in Fig. 1. Seven junction transistors are used.

In the design of the output stage, the desire for low-power consumption and high efficiency led to the choice of class-*B* push-pull operation. The availability of both n-p-n and p-n-p transistors made possible the use of complementary symmetry.[2] Basically, such an arrangement utilizes transistor characteristics whereby the input signal polarity which increases conduction in a p-n-p transistor is opposite to the polarity which increases conduction in an n-p-n transistor. The use of such a circuit eliminates the need for two inputs having opposite phase.

A grounded-base connection is used for output transistors *V1405* and *V1407* be-

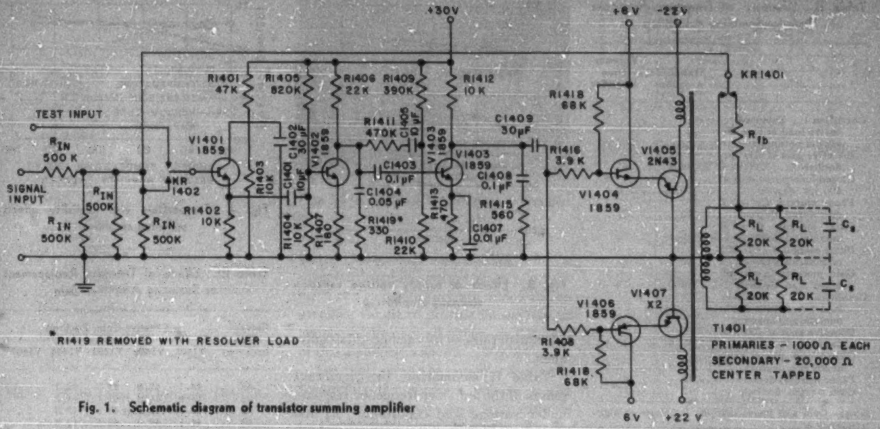
---

Paper 56-261, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing January 20, 1956.

ROBERT C. WEYRICK is with the Goodyear Aircraft Corporation, Akron, Ohio.

The amplifiers described in this paper were developed under contract from Special Devices Center, Office of Naval Research, U. S. Navy. In addition, the author wishes to acknowledge the contributions of Charles B. Barnett.

Vol.
Pt. 1
1956


+30V
+6 V
-22 V
TEST INPUT
SIGNAL INPUT
R IN 500 K
R IN 500K
R IN 500K
R IN 500K
KR 1402
V1401 1859
R1401 47K
R1402 10K
R1403 10K
C1401 10µF
C1402 30µF
R1405 820K
R1404 10K
V1402 1859
R1406 22K
R1407 180
R1409 390K
R1411 470K
C1405 10µF
C1403 0.1µF
C1404 0.05µF
R1419* 330
R1410 22K
V1403 1859
R1412 10K
R1413 470
C1407 0.01µF
C1408 0.1µF
R1415 560
C1409 30µF
R1416 3.9 K
R1418 68K
V1404 1859
V1405 2N43
V1406 1859
V1407 X2
R1408 3.9K
R1418 68K
KR1401
R fb
R L 20K
R L 20K
R L 20K
R L 20K
C s
C s
T1401
PRIMARIES - 1000 Ω EACH
SECONDARY - 20,000 Ω
CENTER TAPPED
6V
+22 V

*R1419 REMOVED WITH RESOLVER LOAD

Fig. 1. Schematic diagram of transistor summing amplifier

cause it exhibits better linearity and d-c stability than a grounded emitter connection. A split-primary output transformer *T1401* is required since the supply voltages for n-p-n and p-n-p transistors are of opposite polarity.

The input resistance of a grounded-base stage is low, generally less than 100 ohms. Because of this low resistance, the output stages are driven by directly coupled, grounded-collector transistors to provide a relatively high input resistance and low output resistance.

With the output transistors operating class *B*, some distortion is present in the output wave form because of increased emitter resistance at low values of emitter current. This distortion is removed by biasing the output transistors so that their operation is slightly class *AB*. These bias resistors are *R1417* and *R1418*. Such operation results in a maximum collector dissipation of approximately 75 milliwatts.

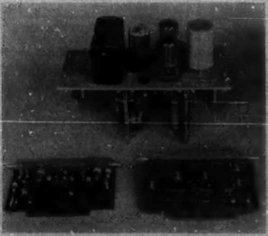


Fig. 2. Two views of transistor summing amplifier (bottom) and its vacuum-tube equivalent (top)

The two types of output transistors used in this stage are Texas Instruments type *X-2* (n-p-n) having a maximum dissipation of 100 milliwatts at 50 degrees centigrade, and Transitron type *2N43* (p-n-p) having a maximum dissipation of 90 milliwatts at 55 degrees centigrade.

The input transistor *V1401* is also wired as a grounded-collector stage to obtain a high value of input resistance at the summing point. This input resistance is approximately 200,000 ohms at 400 cps. Because of the low signal level at this stage, it was considered desirable to reduce the noise level. Since transistor noise increases with collector voltage,[3] the collector voltage of the input transistor was reduced to approximately 3 volts by a divider composed of *R1401* and *R1403*. The voltage gain of this input stage is very near unity. Transistors *V1402* and *V1403* are grounded-emitter stages and provide voltage gains of 17 and 4.5 respectively. The emitter resistors *R1407* and *R1413* provide a measure of gain stabilization.

Because of the large amount of negative feedback around the amplifier, stabilizing networks were necessary to prevent oscillation from excessive phase shifts at high and low frequencies. The high-frequency response of junction transistors is, in general, dependent on both collector capacity and alpha cutoff frequency. Grown-junction n-p-n transistors usually have both a lower collector capacity and a higher alpha cutoff frequency than alloy-type p-n-p transistors. More adequate high-frequency stabilization is therefore obtained from the n-p-n units.

Two phase-lag networks (*R1419*, *C-1404*) and (*R1415*, *C1408*) are included for high-frequency stabilization with a resistive load. The phase margin at high frequencies was further extended by bypassing the emitter resistor *R1413* with an 0.01-microfarad capacitor *C1407*. When the amplifier is used with the tuned resolver load, the network consisting of *R1419* and *C1404* is removed. Low-frequency stabilization is provided by a phase-lead network *R1411* and *C1403* connected between *V1402* and *V1403*.

Relays *K1401* and *K1402* are provided for test purposes. In the operating position, the input signal is connected to the amplifier. In the testing position a known 400-cps test voltage from the trainer is applied to the amplifier, and the feedback resistor is disconnected. This arrangement permits a measurement of the open-loop gain at 400 cps.

The d-c power requirements of the summing amplifier for an output of 125 milliwatts are ±22 volts at 9 milliamperes each, ±6 volts at 9 milliamperes each, and +30 volts at 3 milliamperes, or a total power requirement of approximately 0.6 watt. By comparison the

Table I. Physical Comparison of Transistor and Tube Amplifiers

	Summing Amplifiers (Two Units Each)		Servo Amplifier	
	Transistor	Vacuum Tube	Transistor	Vacuum Tube
Weight, ounces	13	53	51	103
Volume, cubic inches	61	208	98	216

Table II. Summary of Frequency-Response Data for Summing Amplifiers

	Transistor Amplifier	Vacuum-Tube Amplifier
Condition 1: Balanced resistive load of four 20-kilohm potentiometers		
Effective gain at 400 cps, decibels	44.4	45.3
Closed-loop phase shift at 400 cps, degrees	0.4	0.6
Phase margin, C_s* = 0.0 microfarad, degrees	45	51
Gain margin, C_s = 0.0 microfarad, decibels	7.4	15
Phase margin, C_s = 0.002 microfarad, degrees	30	38
Gain margin, C_s = 0.002 microfarad, decibels	9.5	18
Condition 2: Resolver load tuned with 0.27-microfarad capacitor		
Effective gain at 400 cps, decibels	47.6	46.1
Closed-loop phase shift at 400 cps, degrees	0.0	0.0
Phase margin, degrees	41	41
Gain margin, decibels	12	17

Note: Gain and phase margins are those at high-frequency end.

*C_s = shunt capacity across load.

power required for the equivalent vacuum-tube summing amplifier is 8.4 watts.

Different views of two transistor summing amplifiers are shown in the lower part of Fig. 2. Printed circuitry and miniature components were employed throughout. The upper part of Fig. 2 shows a unit containing two equivalent vacuum-tube summing amplifiers. Table I provides a physical comparison of these units. The performance of the transistorized summing amplifier was evaluated by comparing its characteristics with those of its vacuum-tube equivalent.

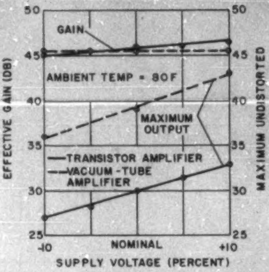


Fig. 3. Effects of supply voltage variation summing amplifier

Table II summarizes the significant points of the amplifier frequency response under various load conditions. At 400 cps the effective gain and the phase shift at a closed-loop gain of 1 are comparable in all uses. The phase and gain margins of the transistor amplifier are, in some cases, less than those of the vacuum-tube amplifier; however, they provide adequate stability against high-frequency oscillation.

Fig. 3 shows the effect of a ±10-percent variation of d-c supply voltage. The maximum undistorted output was determined under closed-loop conditions and a voltage gain of 1.

Fig. 4 shows the effect of ambient-temperature and humidity variations. The effective gain is shown as a function of ambient temperature and relative humidity at an ambient temperature of 90 degrees Fahrenheit. The gain of the transistor amplifier increases with increasing temperature. This gain change develops in the grounded-collector stages *V1404* and *V1406* and is caused by the operating point changing with temperature. This drift also affects the maximum undistorted output of the amplifier at temperature above 120 degrees Fahrenheit. At 130 degrees Fahrenheit, the maximum output

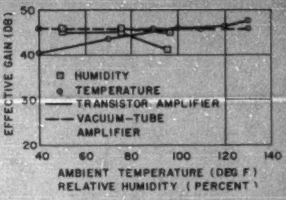


Fig. 4. Temperature and humidity effects summing amplifier

Table III. Effects of Transistor Replacement on Summing Amplifier Gain

Current Gain $\alpha/(1-\alpha)$	Effective Gain, Decibels				
	V1401	V1402	V1403	V1404	V1407
30	44.1	42.8	41.9	42.1	
25	43.8	41.8	41.7	41.7	
38	40.1	42.7	42.8	42.3	
40	44.1	43.0	42.8	42.3	
33	43.5	42.7	42.2	42.2	
46	41.1	42.9	43.0	42.8	
48	42.8	43.2	43.0	43.2	
52	44.5	43.4	43.3	43.3	
56	45.2	42.7	43.1	43.1	
61	44.0	44.0	43.5	43.9	
28					43.5
20					43.4
27					43.7

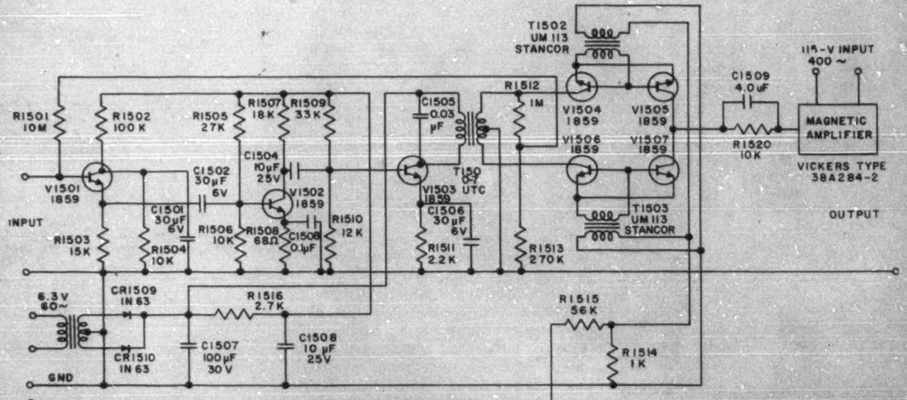

T1502
UM 113
STANCOR
115-V INPUT
400 ~
C1509
4.0 uF
R1512
1M
C1505
0.03 μF
V1504
1859
V1505
1859
V1506
1859
V1507
1859
MAGNETIC AMPLIFIER
R1520
10K
VICKERS TYPE
38A284-2
R1501
10M
R1502
100 K
R1505
27K
R1507
18K
R1509
33K
C1504
10μF
25V
C1502
30μF
6V
V1501
1859
V1502
1859
T150
0-7
V1503 UTC
1859
C1506
30μF
6V
T1503
UM 113
STANCOR
OUTPUT
INPUT
C1501
30μF
6V
R1503
15K
R1504
10K
R1506
10K
R1508
68Ω
C1503
0.1μF
R1510
12 K
R1511
2.2 K
R1513
270K
6.3 V
60~
CR1509
1N 63
R1516
2.7K
R1515
56 K
CR1510
1N 63
C1507
100 μF
30 V
C1508
10 μF
25V
R1514
1K
GND
110 V
400~

Fig. 5. Schematic diagram of transistor servo amplifier

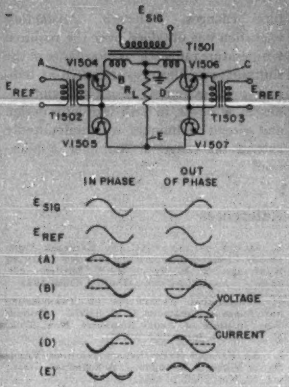


Fig. 6. Wave forms of transistor phase detector

drops from 29 volts rms to 23 volts rms. By comparison, the effects of temperature and humidity over these ranges on the vacuum tube amplifier are slight.

A number of transistors were substituted in the different amplifier stages to determine what effect transistor replacement would have on performance. The gain of the various stages (with the exception of the grounded-base output stages) was found to vary somewhat with the current gain $\alpha/(1-\alpha)$ of the transistor. The other transistor parameters emitter resistance, base resistance, and collector resistance had little effect on gain. Table III shows the effective amplifier gain at 400 cps when transistors having various current gains were substituted in the different stages, and when all voltage amplifier stages except the stage receiving the substitution retained original transistors having approximate current gains of 50 each.

Servo Amplifier

The performance requirements of the servo amplifier are also the same as for its vacuum-tube equivalent and are as follows:

1. The amplifier should be designed to operate a 5-watt servomotor.
2. The amplifier should sum a maximum of four inputs and produce a shaft rotation proportional to within ±1.0 degree of the sum of the inputs.
3. The frequency response of the output shaft should extend to 10 cps.

The transistor servo amplifier consists of a 3-stage voltage amplifier, a phase detector using four transistors, and a magnetic amplifier output stage. A full-wave rectifier and filter are added to the amplifier circuit to provide direct-voltage supply. The over-all schematic diagram of this amplifier is shown in Fig. 5.

Fig. 7. Assembled transistor servo amplifier

The function of the amplifier section is to amplify the resultant voltage of the various signal inputs to a level suitable for the magnetic amplifier input. The input stage is similar to that of the summing amplifier. Transistor *V1501* is connected as a grounded-collector stage to obtain a high input resistance, and the collector voltage is reduced to 5 by a divider composed of *R1502* and *R1504*. Transistor *V1502* is emitter-grounded, and resistance-capacitance-coupled to the input stage. The voltage gain of this stage is approximately 7.

Transistor *V1503* is also a grounded-emitter stage, and its output is transformer-coupled to the phase detector. The maximum undistorted output is 11 volts rms. To maintain this value, the d-c operating point is stabilized by the addition of *R1511* in the emitter circuit. This resistance is by-passed for 400 cps by *C1506*.

Negative feedback is added around the transistor amplifier through resistors *R1512*, *R1513*, and *R1501*. The primary purpose of this feedback is to lower the effective output impedance of the amplifier from about 16,000 to 2,500 ohms. This lower value of output impedance increases the effectiveness of the lead network *R1520* and *C1509* in the magnetic amplifier control winding. This negative feedback, also stabilizes the amplifier gain when transistors are replaced by units having slightly different characteristics.

The function of the phase detector is to convert the a-c output of the transistor amplifier to a polarity-reversing d-c input for the magnetic amplifier. This phase detector employs transistors *V1504*, *V1505*, *V1506*, and *V1507*. The use of transistors for this detector provides two advantages over a diode-bridge type of detector. Since there is no large circulating current through the transistors, no offset or unbalance can develop, and the amount of reference power required is appreciably less.

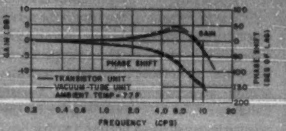


Fig. 8. Frequency response of positional servo system

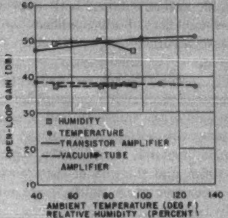


Fig. 9. Temperature and humidity effects, servo amplifier

The operation of the transistor phase detector is shown by the wave forms in Fig. 6. The 400-cps reference voltage is supplied to the emitters through step-down transformers *T1502* and *T1503*. These transformers are phased so that forward emitter current flows in *V1504* and *V1505* during one half-cycle and in *V1506* and *V1507* during the other half-cycle (wave forms *A* and *C*). The collectors of each pair of transistors are connected in series. During a half-cycle of forward emitter current the two paired transistors present a low-resistance path from the secondary winding of *T1501* to the output, and current will flow as shown by wave forms *B* and *D*. During a nonconducting half-cycle, these transistors present a high-impedance path. The resulting output current, wave form *E*, is full-wave rectified d-c and reverses polarity as the signal changes phase with respect to the reference voltages.

Fig. 7 shows the transistor servo amplifier along with the adapter plate used for substituting the amplifier in the computing system. The transistor amplifier and phase detector are on the printed circuit board in the foreground. The magnetic amplifier and its rectifiers are in the rear. Table I physically compares this transistorized unit with its vacuum-tube equivalent. The power required by this transistorized servo amplifier is 0.3 watt as compared to 6.9 watts for its vacuum-tube equivalent.

The performance of the transistor servo amplifier was evaluated while operating in a positional servo system composed of amplifier, servomotor, feedback potentiometer, and summing resistors. A positional servo of this type is used in the computing system to perform multiplication.

The frequency response of this system is shown in Fig. 8 and is seen to be comparable to that obtained with a vacuum-tube amplifier. Both amplifiers showed a static accuracy of approximately ±0.5 degree.

Fig. 9 shows the effect of ambient temperature and humidity on amplifier gain. The gain of the transistor unit is seen to increase with ambient temperature. This change, however, occurs in the magnetic amplifier, and the gain of the transistor amplifier and phase detector are essentially constant.

The substitution of transistors having current gains in the range of 30 to 60 had only a slight effect on amplifier gain. The substitution of transistors in the phase detector had no effect on its operation.

Conclusion

The use of transistor amplifiers in flight simulators as considered herein is both feasible and, from several points of view advantageous. The performance of such transistor amplifiers meets all necessary requirements and is generally equal to that of vacuum-tube equivalents. Furthermore, the transistor amplifiers offer advantages from their substantially lower power requirements, reduced size and weight, and anticipated improvement in reliability. The principal limiting feature of transistor amplifiers is their greater sensitivity to ambient-temperature changes. Although satisfactory operation was obtained over the required temperature range, the performance of the summing amplifier was marginal at temperature extremes. Future improvements, however, in both transistor design and circuit techniques will undoubtedly extend the range of amplifier application.

References

1. DESIGN OF VERSATILE AND RELIABLE COMPUTING COMPONENTS FOR USE IN FLIGHT TRAINING EQUIPMENT, **R. M. Byrne, R. W. Snyder.** *Airborne Electronics Conference*, Dayton, Ohio, 1954.

2. SYMMETRICAL PROPERTIES OF TRANSISTORS AND THEIR APPLICATIONS, **G. C. Sziklai.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, June, 1953, p. 717.

3. AN EXPERIMENTAL INVESTIGATION OF TRANSISTOR NOISE, **E. Keonjian, J. S. Schaffner.** *Ibid.*, vol. 40, Nov. 1952, p. 1456.

4. TRANSISTOR OPERATION: STABILIZATION OF OPERATING POINTS, **R. F. Shea.** *Ibid.*, p. 1435.

Development of Transcontinental Communications in Canada

J. W. NOYES
NONMEMBER AIEE

G. GAUDET
MEMBER AIEE

S. BONNEVILLE
NONMEMBER AIEE

THERE are three major communication agencies in Canada. The Trans-Canada Telephone System furnishes long-distance telephone service to the public across Canada and also provides privately leased telephone and telegraph circuits. The Canadian Pacific Railway Company and the Canadian National Railways each furnish commercial telegraph service throughout Canada, and they also provide a number of privately leased circuits, both telephone and telegraph.

This paper deals with the experience of the Trans-Canada Telephone System in developing a transcontinental telephone network. In pioneering a development of this nature and extent many novel problems were encountered. In addition to the usual design problems, there were others associated with the organization, administration, and operation of an entirely new system across difficult territory. The latter were the hardest to resolve since there had not been any previous experiences of a similar nature from which to benefit, whereas detailed engineering information pertaining to the design of toll plant is obtainable from a number of sources. In view of this it is proposed to discuss mainly the operational problems with only general information on the purely technical aspects.

Geography and Climate

In order to appreciate the difficulties which were encountered in welding this together, some knowledge of conditions in the country being crossed is required.

Canada's enormous area (3,846,000 square miles) makes it the third largest country in the world, exceeded only by Russia and China. It extends from the Atlantic Ocean on the east to the Pacific Ocean on the west; from the Arctic on the north to the United States on the south. While it is a giant geographically, most of the population is within 200 miles of the southern border.

The land consists of the following five geographical areas, as shown in Fig. 1:

1. The Maritime section in the east is part of the Appalachian Region and is in effect an extension of the New England coastal states. It is an area of mountain ranges which have been worn down through the ages. Fishing, agriculture, mining, and lumbering are the main occupations.

2. West of the Maritime Provinces and approximately in the middle of the country is the Laurentian Plateau, a huge U-shaped area of rock, lakes, rivers, forest, and swamp which covers some 45 per cent of the total territory of Canada. It is rich in mineral and forest resources but, with the exception of a few areas, it is not suitable for agricultural purposes. This part of the country is very sparsely populated, even in the south.

3. Further west are the St. Lawrence Lowlands which cover the southern parts of Quebec and Ontario. This is an area of gently sloping fertile plains, comprising about 50,000 square miles. It is the industrial heart of Canada and contains from one half to two thirds of the population of the country.

4. In the middle west is the Great Central Plain, an extension of the midwestern plains of the United States. While this is essentially an agricultural area (grain and livestock), large oil and gas deposits are being discovered.

5. In the extreme west is the Cordilleran or Pacific Coast region, an area, about 400 miles wide, of high mountains, the Rockies, interspersed with valleys. Farming, fruit growing, mining, lumbering, and fishing are the basic industries of this region.

From this description it is obvious that Canada has a difficult topography. Continuous settlement across the country is not possible as the most desirable sec-

---

Paper 56-244, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 20, 1955; made available for printing December 12, 1955.

J. W. NOYES, G. GAUDET, and S. BONNEVILLE are with Bell Telephone Company of Canada, Ltd., Montreal, Que., Canada.

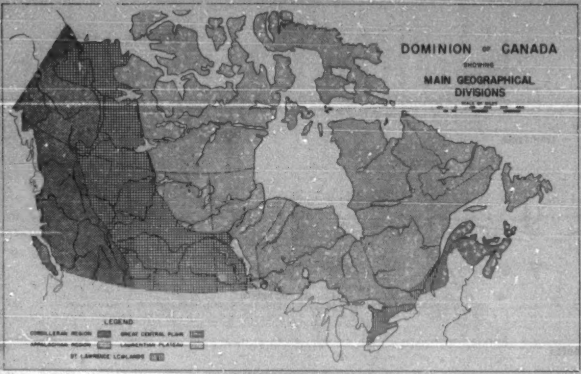


Fig. 1. Main geographical divisions

tions from the point of view of climate and fertility are separated from each other by large areas that can support only a very thin population. This situation is illustrated in Fig. 2 which shows the distribution of population. Two barriers which will be noted particularly are the thinly populated section in the middle of the country and the Rocky Mountain region in the west.

The climatical conditions across the country are difficult also. Hot summers and cold winters give rise to a wide range in temperature, from 60 degrees Fahrenheit below zero to 100 degrees above. Hoarfrost, sleet, and heavy snow are other conditions that have to be considered. The handicaps of topography, geography, climate, great distances, and natural barriers separating populated areas, have posed serious problems in the design, construction, and operation of a transcontinental telephone service.

Background of the System

Telephone service in Canada is provided by seven major telephone systems and together they span the Canadian mainland. Each of them operates an extensive telephone network within its own territory. However, prior to 1928 these systems were fairly well isolated from each other as there were very few intersystem telephone circuits. Most of the long-distance calls between adjacent systems were routed through the United States.

The question of a telephone system across Canada was first studied in 1922. It was considered then that the difficulties in providing a telephone line across the Rocky Mountains, the 1,000-mile stretch of forests, lakes, and swamp from Winnipeg to Sudbury, and the 400 miles of sparsely settled country between the St. Lawrence River and the Maritimes were insurmountable. The matter was not pursued further until 1929 when it was restudied in the light of later technical developments and an expanding economy. The results of this study were favorable, and it was decided to proceed with the construction of a transcontinental telephone system across Canada. To this end the seven major telephone systems serving the country formed the Trans-Canada Telephone System with each one of them providing the facilities needed in its own territory.

An Engineering Committee was formed to co-ordinate the over-all design and construction of the system. It is made up of a representative from each of the member companies, usually the chief engineer. This group is responsible for the development of plans for the construction of additional facilities to meet the forecast for circuit requirements, toll switching plans, and transmission design on both a current and long-range basis. Until recently it was responsible also for traffic forecasts, estimates of circuit requirements, plant maintenance, traffic routing, and other matters.

The route followed across the continent is shown in Fig. 3. It extends from Halifax on east coast to Vancouver on the west, passing through the key cities of St. John, Montreal, Toronto, Fort William, Winnipeg, Regina, and Calgary. The over-all distance is about 4,200 miles.

Each of the nine points is served by long-haul toll circuits which connect it with a number of the other key cities. These circuits constitute the backbone of the Canadian continental telephone network. Each of these offices is a main switching center, and serves as the point of connection with the telephone network in the territory which it serves. For example, all toll traffic with points in the Province of Manitoba and the outside world, is routed through Winnipeg. Similarly, practically all traffic with points in the Province of Saskatchewan is routed via Regina. Exceptions to this arrangement are a few of the other larger toll centers, which have direct circuits to one or more main switching centers in addition to their own home office. The figures show the number of long-haul circuits that were available in 1932 when the

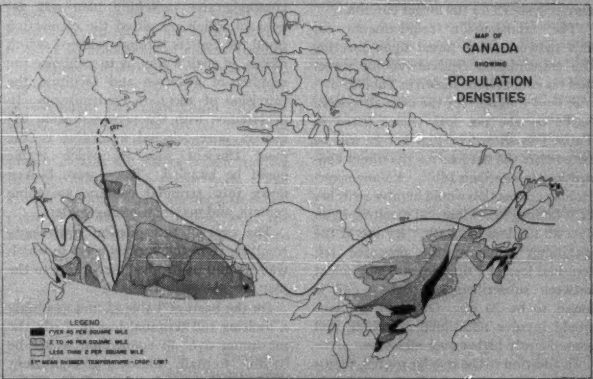


Fig. 2. Population densities

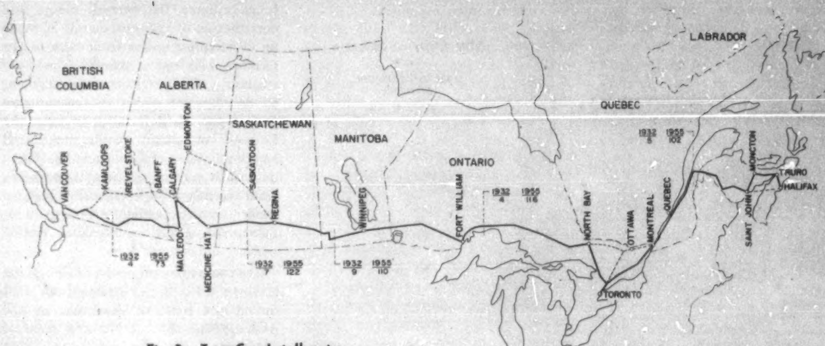


Fig. 3. Trans-Canada toll routes

system was inaugurated and the number that will be available on completion of the current construction program.

Early Design: Open-Wire and C Carrier

The route was developed initially for open-wire and *C* carrier operation. It was designed throughout most of the distance for an ultimate of eight pairs of wire with superimposed *C* carrier systems for long-haul circuits, in addition to the intermediate and short-haul circuits.

TRANSMISSION CONSIDERATIONS

Transmission design requirements in respect to "transmission equivalent," noise, crosstalk, and distortion for both direct and switched connections indicated the desirability of restricting the length of the voice-frequency circuits and using carrier systems for the longer circuits.

The transmission requirements for this network were based upon meeting the transmission equivalents for direct and via connections shown in Table I.

A level diagram for the over-all system based on the net loss values listed in Table I is shown in Fig. 4. It will be noted that the net loss for the direct connections is 9 decibels (db). A Vancouver-Halifax connection would involve switches at Winnipeg and Montreal resulting in an over-all net loss of 15 db. By use of the maximum value of the proposed toll terminal loss of 7 db, the total equivalent between subscribers on such a call is found to be 29 db. Other connections would involve additional switching and consequently higher over-all losses.

In addition to the regular routes, shown for the Trans-Canada System, certain alternate routes shown in Fig. 5 were authorized for service protection. It will be noted that they were provided through connections via United States' points.

At that time two methods were used to introduce gain at switching points, namely, by means of cord circuit repeaters and pad switching. Cord circuit repeaters required the provision of a suitable balanced switching circuit to permit the via circuit to be switched through the repeaters. In addition, suitable switchboard positions were required which involved a 2-jack multiple.

The terminal repeater switching pad plan was just being introduced. This plan provided for the removal of transmission loss at toll switching points by automatic control of artificial lines. This pad plan was used at centers equipped with no. 3 toll switchboards and the cord circuit repeaters were used at others.

As traffic and facilities expanded, a comprehensive switching plan was developed which permitted the connection of all toll centers through selected switch offices, in such a way as to minimize the number of switches and facilitate the provision of adequate transmission. In accordance with the proposed plan, toll centers were selected according to the most efficient way in which traffic could be handled. These were broken down into terminal centers, switching centers, and major switching centers.

The over-all transmission equivalents resulting from the various ways in which traffic could be handled, according to the plan, are shown in Table II.

On the basis of Table II and assuming a limiting toll terminal loss of 7 db, the maximum over-all connection would be 31 db. Actually, the over-all equivalent would vary somewhat from Table II depending upon the minimum working net losses obtainable from the various links. This item was important at that time as open-wire facilities which are subject to weather conditions comprised the majority of the construction.

Because of minimum via net loss considerations it was necessary to use 4-wire circuits equipped with echo suppressors or carrier telephone systems for facilities longer than 1,200 miles. The design of the system covered in general the use of 165-mil copper wire with superposed *C* carrier. With this arrangement it is possible to meet the maximum working net loss limit as suggested. As the system consisted mainly of open wire it was extremely important that the maintenance of the circuits be excellent in order to avoid variations greater than contemplated in setting up the design data. If these variations were exceeded, very serious impairment to transmission would result. Other factors considered at the time were noise, crosstalk, and distortion. The measurement of noise was made on a comparison basis and was expressed in noise units. The noise objective was 200 noise units at the −9-

Table I. Transmission Equivalents for Direct and Via Connections

Connections	1,000-Cycle Transmission Losses (Exclusive of "Toll Terminal Losses"), Decibels
Via circuits for terminal business..........	9
Two via circuits switched together "one switch or two link connections"..........	12
Three via circuits switched together "two circuits or three link connections"	15
Four via circuits switched together "three switch or four link connections"........	18

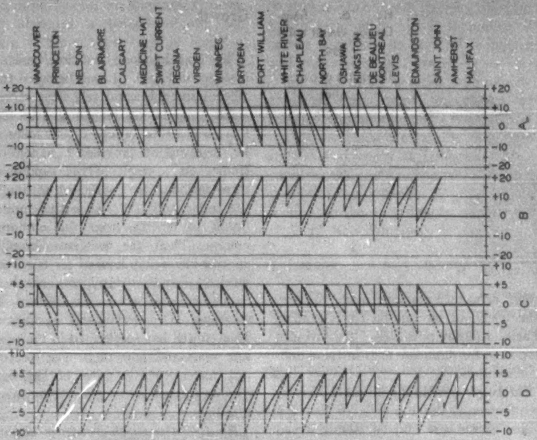


Fig. 4. Transmission levels, representative values

AA—Carrier systems, C type, west to east transmission
BB—Carrier systems, C type, east to west transmission
CC—Voice-frequency open-wire circuits, west to east
DD—Voice-frequency open-wire circuits, east to west, on level diagrams
Solid lines—165-mil copper, 8-inch-spaced CS insulated line facilities (except Oshawa–De Beaujeu 128-mil copper, 12-inch-spaced CS insulated facilities)
Dashed lines—128-mil copper, 8-inch-spaced CS insulated line facilities (except Winnipeg–Sudbury 114-mil copper, 12-inch-spaced CS insulated line facilities)

db level. Crosstalk at that time was measured in crosstalk units and the objective was 1,000 units for carrier and toll cable circuits at the −9-db level point, and 1,500 units for open-wire voice circuits.

Although distortion is a factor, it was not practicable to specify distortion limits except that the design of the component parts of the circuit should be such that distortion impairment was not intolerable. It is of interest to note that at the time it was considered desirable to limit the number of carrier systems operated in tandem to two because of cumulative distortion effects. This resulted in the use of a mixture of voice-frequency open-wire and carrier facilities for long telephone circuits. With present-day knowledge and techniques, such restrictions would not apply.

Where carrier was involved, line filters with a cutoff frequency of 5,000 cycles were used so that the circuits would be available for the transmission of programs.

It was recognized that phantom circuits would have to be abandoned in order to meet crosstalk limits on a transcontinental basis. Although it was known that the limit of 1,500 crosstalk units was somewhat higher than would normally be tolerated, it was felt that it would be a lower limit; in fact it was indicated that in some combinations of closely coupled circuits, somewhat higher values might be experienced. At that time, therefore, development in transmission arrangements envisioned the use of long-haul carrier systems, which eliminated the phantoms and the use of 8-inch spacing between wires of the carrier pairs. The spacing between pairs was increased also from 12 to 16 inches which, with the closer spacing of the wires, operated to reduce the crosstalk effect considerably. The decreased spacing between the wires of a pair also tended appreciably to reduce those components of noise arising from voltage which are directly induced between the wires of telephone circuits by extraneous sources.

New transposition layouts were designed to provide satisfactory noise and crosstalk performance on systems of this length. One of these, an 8-span layout known as the *K8* system, was provided for new construction. A second type of system, based on a 10-span layout known as the *K10* system, was applicable to existing leads on a unit basis without major changes in the existing transposition design.

These transposition systems were being developed while the work was in progress. There was no previous experience with the operation of carrier systems of the length planned for the Canadian network. As a result, a very conservative approach was exercised in the application of the transposition design. Transposition poles were located as precisely as field conditions would permit. This resulted in considerable expenditures which later conditions showed were not necessary.

To obtain suitable transmission results it was necessary to provide the best available grade of outside plant facilities consisting of 165- or 128-mil copper wire on *CS* insulators at 8-inch spacing. The use of these sizes of wire on *CS* insulators has the advantage of lower attenuation and of greater stability compared to the use of smaller gauge wire on ordinary toll insulators. With the larger gauge wire the fluctuations of equivalent are less with temperature and humidity changes. The improved glass insulators, moreover, effected a marked reduction in attenuation and in the variation of attenuation with weather conditions. Eight-inch-spaced nonphantomed facilities were required to meet the crosstalk design limits for circuits of the type and length involved in the Trans-Canada System.

The repeater points were spaced to take advantage of the large-gauge wire

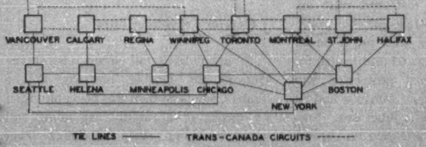


Fig. 5. Emergency traffic layout; alternative routes over long lines facilities

Table II. Over-all Transmission Equivalents

Connections	Transmission Losses, Decibels
Direct circuits, i.e., one link or no switch	9
Two links or one switch connection	8, 10, 11, or 12
Three links or two switch connections	11, 12, 13, or 15
Four links or three switch connections	14 or 16
Five links or four switch connections	17

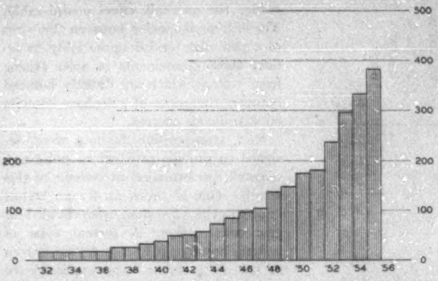


Fig. 6. Trans-Canada circuit mileage

Thousands of circuit-miles

with the improved glass insulators at 8-inch spacing. In establishing the repeater, spacing allowance was made for sleet, hoarfrost, and wet snow conditions. On a call from Halifax to Vancouver, 22 intermediate repeater points were involved. The average repeater section was 180 miles in length with the longest at 250 miles.

Other factors introduced to improve transmission were the use of transposition break irons to provide point-type transpositions. In addition an improved glass insulator mounted on steel pins, bonded together electrically, was developed to improve insulation conditions on the open-wire lines. These were known as *CS* insulators. Subsequent experience showed that the advantages of the bonding were not significant.

Because the Trans-Canada facilities were superimposed on an existing structure, a problem arose in co-ordinating the new requirements of the long-haul circuits with those in existence. In certain cases the repeater points on the two systems did not coincide. The co-ordination of transmission levels was a complex problem and was complicated further by the need for providing emergency carrier assignments over existing 114-mil copper pairs between Sudbury and Winnipeg and 104-mil pairs in other sections.

OUTSIDE PLANT

When the transcontinental toll route was being developed, as much use as possible was made of existing toll leads in the territories of the member systems. To meet the strength requirements of a class *A* line, however, it was necessary to rebuild completely some 2,000 miles of line, equivalent to about half the length of the route, and to add poles in other sections. Where this was done the new poles were spaced on an average 130-foot span basis. In the thinly populated, 1,000-mile section north of Lake Superior, between Sudbury and Winnipeg and through the Rocky Mountains in the west, arrangements were made to carry the telephone circuits on the Canadian Pacific Railway pole lines. This avoided the need for duplicate lines in these difficult sections.

In keeping with the ultimate view of requirements, the top gain, or two top gains, as the case may be, were reserved for the long-haul circuits. As the existing 104- and 114-mil wire on these cross-arms was not considered to be suitable for the backbone circuits from the point of view of strength and transmission characteristics, it was removed and relocated on lower arms. The space thus made available was used for new heavy-gauge wire.

The largest size of standard copper wire for telephone purposes, 165 mil (1/6 inch in diameter), was erected initially over three quarters of the route, with 128-mil copper wire used throughout the balance of the distance. One of the principal reasons for using 165-mil wire was for strength. It was thought that it would withstand the weight of wet snow, sleet, and other mechanical strains better than would the 128-mil wire. Subsequent experience did not support this premise as the incidence of wire breaks on 165- and 128-mil wire did not show any significant difference. It view of this, 128-mil wire has been used in most sections for the construction of additional pairs except on the railway pole line north of Lake Superior where the heavier gauge wire continued to be used.

The entrance cables at the various repeater stations and terminal offices were generally the composite type, with 13-, 16-, and 19-gauge complements, either quadded or paired and loaded as required. Separate east and west entrance cables were installed at each point. The conductors to be used for *C* carrier operation were segregated and loaded for this purpose. Normal protection was provided at the various stations and at the cable and open-wire junctions. Special protection was provided as needed.

Growth

The growth in long-haul circuit mileage is shown in Fig. 6. It will be noted that when transcontinental telephone service across Canada was inaugurated in 1932 there was a total of only 20,000 miles on long-haul circuits provided to handle the traffic. As a result of the depression, these facilities were sufficient to handle the traffic until 1937 when the country began to emerge from the economic morass. Then, the war years brought an ever-increasing demand for telephone service, as indicated by the growth in circuit mileage which was limited by shortages of material and equipment. This growth has accelerated in the post-war years and now there is a total of just under 400,000 miles of long-haul circuits carried across Canada on open-wire lines.

By 1947 the rapid growth in circuits needed to handle the traffic began to exhaust the *C* carrier capacity of the open-wire line. About 30,000 miles of long *C* systems, or 90,000 channel-miles, were in operation across the country, of which almost 60 per cent were made up of systems 1,000 miles or more in length. Some typical systems are Toronto-Calgary, 2,170 miles; Montreal-Winnipeg, 1,400 miles; Toronto-Winnipeg, 1,250 miles; and Regina-Vancouver, 1,130 miles.

The next step in the race with the rapidly expanding demand for telephone service across the continent was develop the open-wire line for *J* carrier operation. The first *J* system was placed in service between Toronto and Winnipeg in 1948. The growth in the development of broad-band open-wire carrier systems across the country has been rapid. There are now over 16,000 miles of *J* systems in service providing a total of about 200,000 circuit-miles, or about half of the total long-haul circuit mileage across the country. Systems over 1,000 miles in length include Toronto-Winnipeg, 1,250 miles; Montreal-Winnipeg, 1,400 miles; and Winnipeg-Vancouver, 1,700 miles. The *J* systems have been used for the longer circuits with some of the *C* systems cut up to provide circuits between adjacent switching centers, e.g., Calgary-Regina and Regina-Winnipeg circuits.

J Carrier Developments

By about 1945, toll circuit relief in the Ontario and Quebec portions of the Trans-Canada System was being provided by means of a rapidly expanding toll cable backbone route extending from

Toronto to Quebec via Kingston, Ottawa, and Montreal, a distance of over 500 miles, with a network of cable branches reaching most of the important traffic centers in the territory. Voice-frequency loaded cable facilities and short-haul carrier systems were assigned to relatively short distance circuits, while *K* carrier facilities were assigned to the longer circuits in the backbone route.

For the remaining, and by far the longer, portion of the Trans-Canada route west of Toronto and east of Quebec, engineering studies indicated unmistakably that an extended use of existing open-wire facilities by means of *J* carrier offered the most practical means of relief for the time being. This plan, however, presented some unusual technical difficulties which will be discussed in the following.

Introduction of *J* carrier operation made it necessary to carry out extensive changes and rearrangements to the open-wire lead to make it suitable for this type of operation. This included retransposition, the making of gaps of at least 80 feet in the open-wire line at each repeater station, installation of new entrance cables and rearrangement of existing cables, installation of suitable crosstalk suppression and protective devices on each pair of wires at the terminal poles, construction of new auxiliary repeater buildings, etc.

The most critical bottleneck was between Toronto and Winnipeg, 1,250 miles, largely because of the section Sudbury-Winnipeg, 978 miles, where the Trans-Canada facilities consisted basically of four pairs of 165-mil copper wire on the top arm having an 8-inch–16-inch–8-inch configuration, with *K10* transpositions and full *C* carrier development. These 165-mil top arm pairs were in the somewhat disturbing company (from a transmission co-ordination standpoint) of non-co-ordinating Railway Company 114-mil pairs on lower arms and frogged steel and copper-pole pairs of wires on the two top arms. To increase the complexity of the problem, in many sections the lowest Railway Company crossarm carried "untouchable" block signal wires which were outside the authority of the Railway Company's Communication Department. It was anticipated that these wires might act as tertiary paths which could contribute to run-around crosstalk or singing.

Because this Toronto-Winnipeg section was the first section in the Trans-Canada route where *J* carrier was developed, and because it also presented some of the most challenging problems, the treatment chosen will be described as typical of problems met in subsequent *J* carrier developments west of Winnipeg and east of Quebec.

Since *K* carrier facilities were already available from Toronto to Oshawa, an obvious solution was that *J* systems west of Oshawa should be extended to Toronto by *J-K* conversion at Oshawa. The *J* carrier design problem was now restricted chiefly to the Canadian Pacific pole-line section in the rugged country between Sudbury and Winnipeg.

J REPEATER SPACINGS

In the layout design of a *J* carrier system of more than 1,000 miles in length, the choice of repeater spacings (and thereby the determination of the number of repeater stations) has a major influence on the cost of the project and on the expected performance.

It is an inherent weakness of open-wire carrier systems that they cannot be built economically for complete immunity to service interruptions under extreme weather conditions. A compromise has to be made at some point where greater insurance against service interruptions would necessitate disproportionate increases in costs. In this project, it was decided to use a design aimed at limiting system outages to not more than 5 to 10 days per year.

A program of high-frequency transmission-line loss measurements was undertaken for several strategically chosen sections between Oshawa and Winnipeg to indicate the range of line-loss variations that could be expected in different regions along the route and over the winter months. Inquiries were also made of meteorological bureaus, farmers, and railway and telephone craftsmen, as to their records or recollections of sleet, freezing rain, hoarfrost, or wet snow incidence in different line sections. Information from such sources provided a surprising amount of contradiction but was of some assistance.

The exact choice of repeater station locations was influenced by many factors. It was important that each station be close to an existing railway station or siding and that it be at or near a neutral point in transposition layout (*S* point) to minimize the need for retransposing. Terrain had to be considered. For instance, marshy ground or steep slopes were not acceptable. Availability of power supply was another factor.

After all the afore-mentioned aspects had been duly considered, a *J* repeater layout was designed with spacing ranging from 35 to 70 miles, and transmission capability of overcoming open-wire losses under ice conditions of more than 2.0 db per mile in the sections considered most exposed to ice incidence down to as little as 1.1 db per mile in sections estimated as relatively immune.

ELECTRICAL CO-ORDINATION

The initial project of *J* development in the bottleneck, Sudbury-Winnipeg, involved only one system in the first stage, to be followed by a second system some 2 years later, and with an ultimate design of eight or possibly nine systems.

Attenuation-frequency measurements showed that the existing top arm with *K10* transpositions had two pairs on pins 1–2 and 7–8, free from attenuation peaks and suitable for *J* carrier. Of course, the design of the first system involved no *J* crosstalk problem, but it did present unusual repeater singing problems which, without proper remedial measures, would have reduced the usuable maximum repeater gains of 77 db by an intolerable margin of 15 db or more.

At auxiliary repeater stations, standard suppression measures were estimated to be adequate, and this was confirmed by later experience. The railway signal wires on bottom crossarms were a cause of worry but actually gave little trouble.

The chief difficulties were found at some of the main repeater stations, which already served as *C* carrier and voice-frequency repeater points, where loaded entrance cables were unusually short and did not provide the degree of *J* carrier frequency crosstalk and singing suppression (via tertiary paths) as is normally assumed for main repeater stations. In one case, the problem was complicated by the adjacency on both sides of the repeater station of another railway pole line which acted as a tertiary path.

These abnormal situations were controlled by unusual application of filtering arrangements normally reserved for auxiliary stations. These in turn necessitated special bay layout arrangements, and in one case building modifications to increase ceiling height beyond what would normally have been required.

CONSTRUCTION PROBLEMS

Most of the proposed auxiliary *J* repeater stations were to be in wild bush country, accessible only by railroad. Arrangements were made with the Railway Company to build sidings where necessary, to provide suitable railroad cars for the living accommodation of the installing crews, and to serve as storage sheds for equipment. Portable oil heaters and gasoline-driven electric generators for

Fig. 7 (above). Auxiliary J repeater building

Fig. 8 (right). Bunkhouse

lighting and for auxiliary electric service supplies (such as soldering irons, etc.) had to be provided. In several cases, automatic gasoline power plants had to be installed and suitable arrangements had to be made for gasoline storage which conformed to local regulations. Ventilation, for summer and winter, of unattended auxiliary repeater stations with automatic power plants presented many problems; for instance, poisonous exhaust gases from the gasoline engines being sucked back into the engine room by the automatic intake fans. There was also a danger of the exhaust pipe of the engines which projected outside of the building becoming choked by snow drifts or icicles, thereby incapacitating the engines. However, these problems, as well as many minor ones, were finally solved.

TRANSMISSION DESIGN FOR THE SECOND AND ADDITIONAL J SYSTEMS

A second *J* system could have been installed on pins 7–8 between Sudbury and Winnipeg while the first system was operating on pins 1–2 without changing the existing *K10* transpositions. With a view to the future, it was realized that the next step would involve retransposition to some suitable *J* design and that execution of this plan would result in many interruptions or impairments of service while the retransposition was underway. It was therefore decided to retranspose the top arm to the *J5* design before the second *J* system was installed.

At this stage, the problem of crosstalk between *J* systems arose. Because of the great length of the systems, and of the many adverse conditions involved, it was decided to take other measures to insure control of crosstalk beyond retransposing the top arm. These measures included the following:

1. Respacing of pins on top crossarm from the 8–16–8-inch configuration to 8–24–8-inch configuration. Six-inch wire spacing was considered unsuitable because of the liability of wire contacts under prevailing weather conditions. The respacing from 8–16–8- to 8–24–8-inch configuration involved a question as to whether old crossarms, originally bored for 12–12–12-inch configuration, then rebored for 8–16–8-inch spacing, and now to be re-rebored on 8–24–8-inch basis, could take the reboring without dangerous weakening. Studies and tests led to the conclusion that if the crossarms were free from decay, the reboring should cause no weakening as the critical breaking point was close to the pole and was not affected by reboring along the arm.
2. Removal of pole pairs on two top arms. This was predicated by step 1, due to space considerations, and because in any case the *J5* transposition does not provide for pole pairs.

FINAL TRANSMISSION DESIGN

With repeater singing and run-around crosstalk assumed to be under control, there remained two outstanding problems: noise and echoes.

These challenges were met by the use of compandors on critical channels, and by the judicious application of echo suppressors.

UNUSUAL PROTECTION PROBLEMS

The western portion of the Trans-Canada route is laid out roughly along the 50 degree parallel of North latitude. This geographical situation, in combination with the prevalent ground conductivity, may account for the severe electrical disturbances on line wires which have been experienced, particularly between Sudbury and Winnipeg, as a result of magnetic storms. Sporadic periods of rapidly fluctuating voltages between line wires and ground with peaks of between 500 to 1,000 volts have been experienced year after year. In these magnetic storm periods, carbon protectors broke down and some remained permanently grounded which resulted in circuit unbalances or short circuits. For stations in bush country with access difficulties, these protector troubles often resulted in seriously long service interruptions.

To remedy this situation, drainage coils were installed in association with carbon protectors, connected in such a way that permanent grounding of a pair of carbon blocks would affect only direct currents and would leave all carrier frequency currents relatively unaffected. Installations of 1,000-volt protectors were also made at strategic points to minimize chances of service interruptions due to troubles at points not readily accessible.

OUTSIDE PLANT PROBLEMS

At main repeater stations where long entrance cable facilities were already in place, pole-mounted *J* line filter installations were provided to separate the *J* carrier frequency currents and transmit them over nonloaded paper-insulated entrance cable pairs. Unfortunately, unlike so-called "junction filters" which separate all carrier from voice frequencies, these *J* line filters are sensitive to low temperatures and are liable to be permanently damaged if exposed to temperatures below freezing. Special pole-mounted filter boxes were therefore designed with thermostatic control of electric heating units and automatic alarms in the near toll office when temperature in the filter box dropped to near-danger point. Provision was also made at each box to supply emergency power for the heaters from the repeater station 130-volt battery in the event of commercial power failure.

Vol.
Pt. 1
1956

Entrance cables at the various stations were generally underground. In these locations where the entrances were in muskeg or low lying land and where drainage was impracticable, paper-insulated lead-covered cables were placed in transite ducts, not encased, and the spiral four cables were placed inside $1^{1}/_{2}$-inch-diameter steel pipe. When the pipe joints were properly made and the pole end was adequately sealed, no trouble has been experienced. However, where the joints or seals were not properly made, water entered the pipes and during the winter caused spiral four cable failures due to crushing by ice in the pipe.

Span-type transposition brackets were used in connection with the retransposing to avoid extensive pole respacing, although there had been very little actual experience with this type of bracket at the time. The brackets were made of magnesium initially to keep their weight down to a minimum. This did not prove to be satisfactory, however, because of frequent failures due to fatigue. To overcome this, aluminum has been used for new brackets with excellent results. Later experience has raised doubt as to the efficiency of span brackets in reducing crosstalk under some conditions.

J Carrier Development West of Winnipeg and East of Quebec

The foregoing outline of the engineering design of the *J* carrier systems between Toronto and Winnipeg covers all the major problems that had to be met on other Trans-Canada *J* carrier developments. The prairies presented problems of access to stations and of the possibility of heavy hoarfrost in several sections and frequent gale force winds in other sections, as well as an extreme range of temperatures. In Alberta temperatures sometimes vary by 50 degrees in less than 1 hour when the Chinooks blow in from the Rocky Mountains.

The *J* carrier development in British Columbia naturally presented a severe engineering problem because of the extremely mountainous nature of the country, the wide weather variations, and the snow and rock slides. In the section in between Hope and Princeton, the Trans-Canada wires were on a railway pole line traversing a valley where snow slides had several times carried away miles of pole line and had buried railway tracks. To protect the *J* carrier systems against this possibility, a new pole line was built along a safer route. This line was provided initially with three pairs of 8-inch-spaced 128-mil copperweld wire, *J5* transposed.

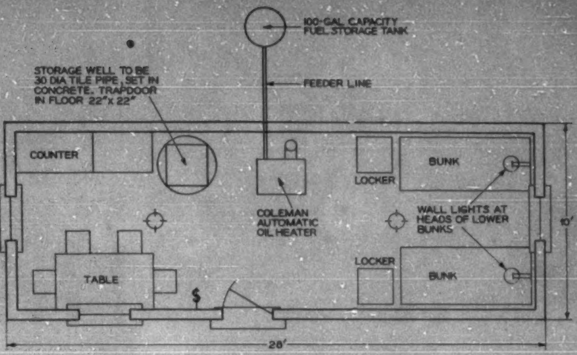

100-GAL CAPACITY
FUEL STORAGE TANK
STORAGE WELL TO BE 30 DIA TILE PIPE, SET IN CONCRETE. TRAPDOOR IN FLOOR 22"x 22"
FEEDER LINE
COUNTER
BUNK
LOCKER
COLEMAN AUTOMATIC OIL HEATER
WALL LIGHTS AT HEADS OF LOWER BUNKS
10'
TABLE
$
LOCKER
BUNK
28'

Fig. 9. Plan of bunkhouse

Arrangements with Railway Company

As indicated earlier, the 1,000-mile, thinly populated section north of Lake Superior was a serious obstacle to the development of a transcontinental telephone service across Canada. Construction of a pole line across this barrier was out of the question at the time. The cost would have been prohibitive, maintenance would have been very expensive, and the volume of business to be handled was relatively light. Indications were that it would be many years before the volume of traffic could carry the cost of an open-wire telephone line.

The Canadian Pacific Railway had a pole line in this section which was built in conjunction with the railway line. It was used for telegraph business and for railway operations. By arranging with the Railway Company to carry the telephone circuits on its pole line, it was possible to bridge this gap.

The basic agreement with the Railway Company covers the construction of four pairs of wires on the top crossarm over a distance of 1,000 miles for telephone purposes, including entrance cable facilities. They were constructed to the Telephone Company's specifications and leased to them exclusively for an initial period of 20 years which has subsequently been extended. The Telephone Company was given the right to operate at any frequency above 100 cycles per second over them, and the railway used the band under 100 cycles for d-c telegraph service. The operation of any form of commercial telegraph over these facilities by the Telephone Company was expressly forbidden.

The cost of rearrangements to the railway plant on the line to provide space for the new pairs and the cost of retransposing existing railway facilities to coordinate with those constructed for the use of the Telephone Company was borne by the latter.

The Telephone Company provided its own *C* carrier and voice-frequency repeaters, test equipment, etc. It was located in space made available in the Railway Company's existing repeater offices on the route. The Railway Company maintained the telephone equipment in its offices and also furnished power.

With the introduction of *J* carrier on the lead by the Telephone Company, it was necessary to make extensive changes and rearrangements to the pole line to make it suitable for this type of operation. The Railway Company carried out all the rearrangements and changes and built the new buildings to the specifications and plans furnished by the Telephone Company. The Telephone Company reimbursed the Railway Company for the cost of the retransposition, the changes, and the rearrangements, and leased the new buildings.

Rental Rates

Open Wire

An interesting feature of the arrangements with the Railway Company is the formula developed to determine the rental rates for the pairs of wires leased to the Telephone Company. These pairs were constructed at different times, and the formula was designed to reflect any variations which might occur in the price of copper during this period. It served to protect each of the contracting parties as the rental charge for a new pair of wires could be determined within close limits before the work was authorized.

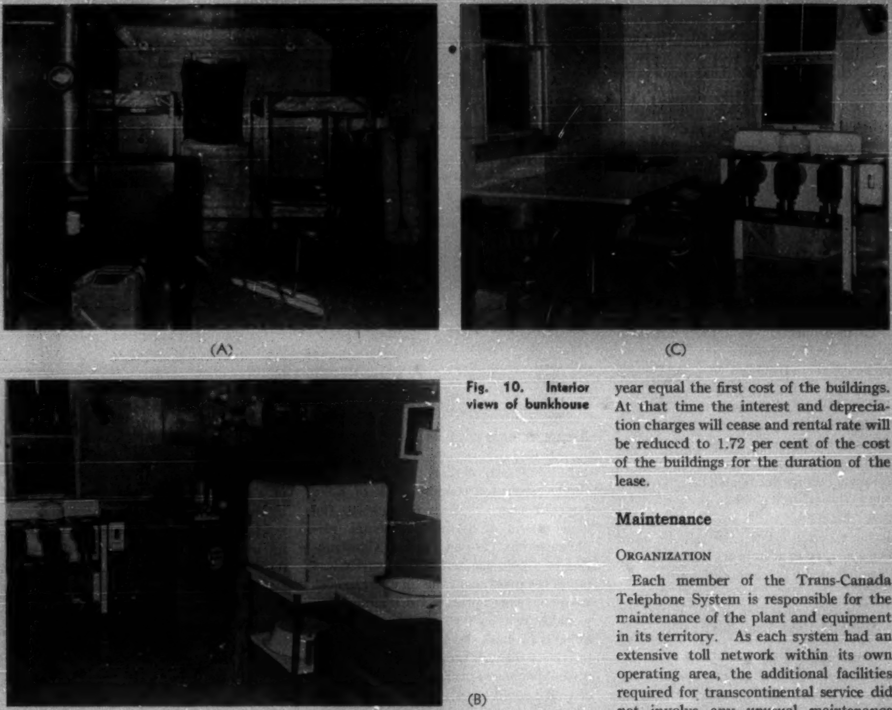

(A) (C) (B)

Fig. 10. Interior views of bunkhouse

The general formula is as follows:

$\$28.43+(1.072\times N)$ where N is the cost of copper wire in cents per pound

The formula is based on costs which prevailed before World War II and a route mileage of 973 miles.

In the case of the first pair of wires provided for the use of the Telephone Company, the rental rate derived by means of the formula was increased by a fixed amount per circuit-mile per year. This amount covers the charges for floor space reserved for the Telephone Company's equipment in the four Railway Company main repeater stations; house service, insurance and taxes on this space, replacement of crossarms, entrance cables, and cable loading. At such time as the lease for the pair carrying this charge expires, the amount is then added to the rental for one of the other pairs.

Auxiliary J Repeater Buildings

As indicated previously, the Railway Company constructed the new buildings required at the auxiliary *J* repeater sites and leased them to the Telephone Company. The buildings are located on the railway right-of-way at locations selected by Telephone Company engineers. The heating and ventilating equipment, power plants, fuel oil storage, and gasoline storage where required, as well as all the carrier equipment and associated apparatus at the auxiliary repeater stations, was provided by the Telephone Company.

The rental for these buildings is based on the exact cost of the buildings and includes the following items:

	Per Cent Per Year
Interest	5.0
Depreciation	3.27
Administration	0.5
Maintenance	0.5
Taxes	0.37
Insurance	0.10
Use of right-of-way	0.25
Total	9.99

This rate is to be used for determining the rental charges until the amortization charges accumulated at 5.0 per cent per year equal the first cost of the buildings. At that time the interest and depreciation charges will cease and rental rate will be reduced to 1.72 per cent of the cost of the buildings for the duration of the lease.

Maintenance

ORGANIZATION

Each member of the Trans-Canada Telephone System is responsible for the maintenance of the plant and equipment in its territory. As each system had an extensive toll network within its own operating area, the additional facilities required for transcontinental service did not involve any unusual maintenance difficulties. The main problem was that of training the plant forces in the special problems associated with the introduction of new instrumentalities, techniques, and operating procedures.

The most difficult maintenance problem concerned the facilities carried over the railway pole line north of Lake Superior. During the early years when the facilities consisted of open-wire and *C* carrier, all the repeaters were located in the Railway Company offices. These offices were staffed by railway employees trained to maintain and operate the communication facilities of the railway. Arrangements were made to have this staff maintain the telephone equipment also. To this end the men were trained in Bell System techniques and methods, and furnished with the necessary test gear and practices. Their salaries were prorated between the Railway Company and the Telephone Company in proportion to the work performed for each company. This is determined by periodic studies involving an inventory of the equipment owned by each company in each of the main repeater offices and an estimate of the man-hours required for both equipment and service maintenance on each company's equipment using work units derived from Bell System experience.

Introduction of *J* carrier operation across the continent gave rise to new maintenance problems, such as the maintenance of terminal and repeater equipment at locations where a regular maintenance staff is located. The problems here are quite similar to those encountered in maintaining other types of toll equipment. Time spent on routine maintenance of the *J* equipment in the attended stations is not excessive, amounting to about 6 hours per quarter per terminal and 3 hours for each repeater.

Periodic interchange of located personnel with those assigned to the maintenance of auxiliary stations on a despatch basis is desirable. This scheme results in periods of greater supervision and training. It also permits greater flexibility in the use of available staff.

Servicing of the auxiliary *J* repeater stations is done on a motorized despatch basis except in the undeveloped section north of Lake Superior. In a reasonably well developed territory, regular toll offices are distributed so as to permit the despatch of personnel from these offices to the adjacent auxiliary stations.

Routine maintenance requires that each auxiliary station be visited every quarter. While the number of working systems in a station governs to some extent the maintenance effort required, experience shows that these visits consume about 50 hours per station per quarter. Travel time is in addition to this and varies according to the distance that must be travelled. Visits in connection with trouble alarm indications have been averaging about one per auxiliary station per month. Time taken on these visits varies with the nature of the trouble with travel time, of course, in addition.

Maintenance of the unattended *J* repeater stations in the undeveloped section along the railway line north of Lake Superior presented a difficult problem. Most of these stations are not accessible by road and can only be reached by train, usually freight.

Two plans for maintaining these offices were studied. One was to use Telephone Company personnel, and the other was to have the Railway Company do the work. The latter plan was adopted as the most satisfactory solution. Arrangements were made to have three railway telegraph men assigned full time to take care of the equipment at the auxiliary *J* repeater stations. The Railway Company is reimbursed for the wages and expenses of these men by the Telephone Company. Each man has four to five stations assigned specifically to him. He uses the main repeater station in his section as headquarters.

At nine of these stations accommodation for lodging and meals does not exist as they are miles from any habitation. Fig. 7 shows a repeater building with a bunkhouse. The dead-end pole on one side of the station can be seen also. The wires on the other pole are signal wires which the Railway Company would not allow to be carried in and out of the repeater station. As the train schedule does not permit a man to enter and leave one of these stations on the same day and still have adequate time for maintenance work, it was necessary to provide some form of living accommodation. The first approach to the problem was to provide a folding cot, blankets, electric hot plate, dishes, cutlery, and cooking utensils in the repeater station but this arrangement was far from satisfactory. Meals were prepared under difficulty with little space available for eating and storage of food or utensils. Sleeping among bays of clicking relays and whirring motors was not conducive to a good night's rest, and the depressing influence of battery fumes, absence of natural light, and cold floors added to the inhospitable living conditions.

The solution to this problem was the construction of wooden bunkhouses, 28 by 10 feet in size, adjacent to the auxiliary repeater station, see Fig. 8. The buildings are insulated and lined on the inside with plywood. Fig. 9 shows the floor plan and arrangement of the different items of equipment. Interior views of these cabins are shown in Figs. 10(A), (B), (C). It will be noted that the buildings are well furnished with an oil heater, oil stove for cooking, two double-deck bunks, a small electric refrigerator, table, chairs, kitchen utensils, and blankets, and that they are stocked with provisions.

The buildings were built by the Railway Company and are leased to the Telephone Company. The rental paid is based on a formula similar in principle to that used in determining the rental charges for the auxiliary repeater buildings.

A number of auxiliary repeater stations where commercial power was not available were equipped with gasoline-powered engine-driven generators. The maintenance of the engines was kept up by the Telephone Company and not the roving attendants. These stations were visited every 2 months in winter and 3 months in summer by two mechanics from the motor vehicles section who spent 3 days in each station checking the engines.

Table III. Plant Maintenance Results

Cross Section	First Six Months 1954	First Six Months 1955
Montreal and Toronto-Maritimes	5.52	5.74
Montreal and Toronto-Winnipeg	8.58	4.25
Winnipeg-Regina	9.47	7.48
Regina-Calgary	6.86	11.83
Calgary-Vancouver	15.95	10.84

Several years' experience with engine-operated power plants has resulted in a number of modifications to improve their over-all operation. Two of the most important changes are as follows:

1. The normal ventilating arrangement of the engine crankcase provides for feeding the exhaust oil fumes back into the intake manifold. The oil vapor forms carbon when it comes in contact with the hot intake valves. Accumulation of carbon caused the valves to stick and burn which necessitated a regrinding job every 3 months. To overcome this, each engine was modified to exhaust the fumes directly outdoors. This has reduced the frequency of the valve-grinding job.
2. The oil capacity of the engine crankcase is 5 quarts. The oil consumption required a visit to each station every 3 weeks to replenish the oil. The need for these frequent trips was overcome by installing a 5-gallon reserve oil supply tank which fed oil into the engine crankcase under control of a float valve.

It should be pointed out that power supplied by engine-operated power plants has proved to be a costly undertaking and should be avoided if it is at all possible to bring in commercial power.

RESULTS

The maintenance results experienced on this network follow the general pattern for open-wire plant. The principal sources of trouble have been due to sleet, hoarfrost, wet snow, protector operation due to voltage surges, broken insulator, swinging short circuits, train wrecks, highway accidents, etc. Particular mention should be made of snow slides and rock slides in the Rocky Mountain section which took the lead out for days at a time. This condition was improved by a reroute of the line in the worst section coincident with the *J* carrier installation. It is also interesting to note that interruptions due to hoarfrost or wet snow have mostly occurred during early morning hours, when traffic was light, and clearance generally happened soon after sunrise.

The over-all performance of the carrier systems has been excellent with no unusual maintenance and operational problems arising out of their extreme length. While several of the systems are approximately 1,000 miles in length, it was recognized that this was about the upper limit for satisfactory regulation. Consequently, the 1,700-mile Winnipeg-Vancouver *J* system now being installed will have its regulation broken at Calgary into two sections.

Recent plant maintenance results are listed in Table III. They show the lost time per 100 circuit-miles for the Trans-Canada Telephone System facilities from 9 a.m. to 10 p.m. Previous results are not directly comparable. As the first 6 months include the severe winter months, the results for them are much worse than for the balance of the year.

Future Development

The transcontinental open-wire lead has been developed to its maximum capacity. It is of interest to note that the *J* carrier capacity of the line has been reached only 8 years after the first broad-band carrier system was placed in service. The high rates of traffic growth now being experienced and the need for television facilities requires a new basic structure.

Long-term relief is being provided by a coast-to-coast radio-relay system carrying television and *L* carrier facilities. This system is now under construction with over-all completion scheduled for late 1957 or early 1958. A light-route microwave system installed earlier between Halifax and St. John for telephone service has been operating satisfactorily. It will be noted that the coaxial stage of development used in the United States has been avoided in Canada. Because of the timing, Canada has gone directly from open-wire to radio-relay operation.

The additional telephone circuits required for direct distance dialing (customer toll dialing with automatic machine switching) will be furnished economically over the new radio-relay structure. This system will also tie in with the new transatlantic cable.

Design of Electronic Equipment for Radio-Interference Reduction

A. L. ALBIN
ASSOCIATE MEMBER AIEE

H. M. SACHS
NONMEMBER AIEE

Synopsis: During a survey recently conducted for the U. S. Army, Navy, and Air Force, it was found that the control of radio interference was a serious problem to both designers of electronic equipment and to the military operational activities. Some of the critical areas of design that have affected equipment and system performance in the field are discussed herein. Several types of electronic equipment for ground and air-borne applications are discussed including radar, communications, and navigation apparatus. The effectiveness of practical suppression techniques such as shielding and filtering is also considered. Requirements of the current interference specifications are reviewed and techniques discussed for making conducted and radiated measurements of continuous-wave and broad-band interference sources.

THE Armour Research Foundation recently completed a study and evaluation of the radio-interference program of the services. The purpose of the study was to make a comprehensive survey and formulate recommendations for a long-range program of research and development, and to recommend measures which could be put into effect immediately in such areas as maintenance and operation of equipments in the field.

The survey was made by investigations conducted at military operational activities, industrial organizations, and government research laboratories. Interviews with engineers and operating personnel at these facilities indicated that, while many of the problems were operational in nature, others were caused by a lack of consideration of radio interference in the initial design of electric and electronic equipment. Much technical information is now available on procedures for interference-free design, and the services have made great strides in their efforts to awaken an interference consciousness in engineers not normally concerned with these problems. It was found, however, that organizations not ordinarily concerned with the radio-engineering aspects of interference, such as manufacturers of ships, electric motors, and vehicles, desired more information of a tutorial nature to acquaint them with applicable design, development, and measuring techniques. It is to this group that this paper is directed.

This paper will present some of the general design problems that are encountered, remedial measures that have proved effective in practice, and a brief discussion of applicable measurement techniques. In addition, the references will be found to provide valuable design information for specific applications.

General Suppression Techniques

NEED FOR EARLY CONSIDERATION

For effective application of radio-interference suppression, it is important to consider the problem early in the design stage. While considerable interference reduction can often be achieved by conservative design, it is frequently found that additional components must be incorporated in the equipment to obtain the desired degree of interference reduction. Consequently, it is of advantage to anticipate possible difficulties and to provide room in the equipment for such components. Furthermore, conventional parts may often be replaced by others with superior interference-reduction properties at little or no additional cost, provided the designer is aware of their application and availability.

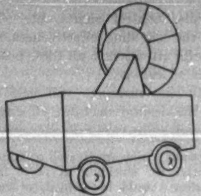
SOURCES OF INTERFERENCE

Undesired coupling leading to radio interference may occur by radiation, induction, or conduction. Since so-called radiated measurements are often made in the induction field, it has become general practice to include the influence of both the near and far fields within the scope of radiated interference. Although not

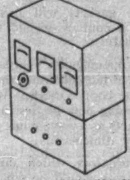
---

Paper 56-257, recommended by the AIEE Special Communications Applications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N.Y., January 30–February 3, 1956. Manuscript submitted October 17, 1955; made available for printing December 16, 1955.

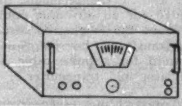
A. L. ALBIN and H. M. SACHS are with the Armour Research Foundation, Illinois Institute of Technology, Chicago, Ill.



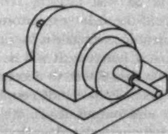
(A) RADAR



(B) COMMUNICATIONS TRANSMITTER



(C) SUPERHETERODYNE RECEIVER



(D) ELECTRICAL DEVICES

Fig. 1. Interference sources

strictly a rigorous interpretation, it serves as a practical distinction from the mode of transmission wherein coupling occurs by means of currents propagated along circuit conductors or ground paths. Fig. 1 illustrates representative interference sources, while Fig. 2 shows the aforementioned methods of interference propagation.

While interference generators may produce wave shapes of many varieties, those of interest herein may be grouped broadly into two general areas: broad-band and continuous-wave sources; see Fig. 3. Broad-band interference may be produced by pulses of short rise and decay times and narrow widths, which have energy content extending over a wide range of frequencies. A radar modulator, for example, produces a pulsed output signal at a fundamental repetition rate of 400 to 3,000 cycles per second (cps), yet may contain significant energy components as high as 30 megacycles. Continuous-wave interference is that produced at discrete frequencies because of such effects as harmonics, parasitic oscillations, and other generators of primarily sinusoidal waveform.

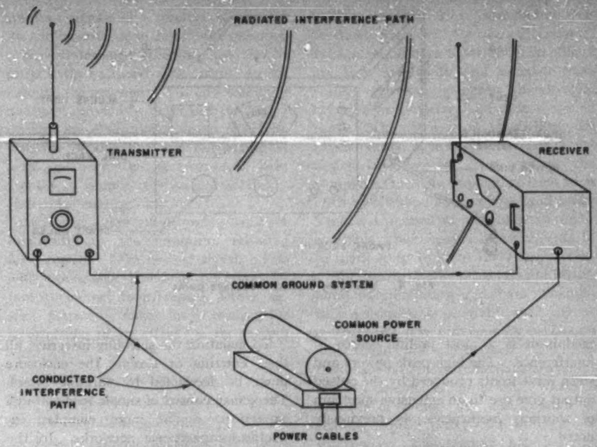
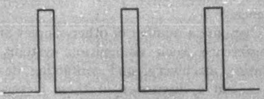


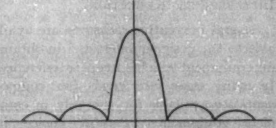
Fig. 2. Conducted and radiated interference propagation

Almost any radio transmitter can be a source of interference unless extreme care is given to control the generation of spurious emission. The need for accurate frequency control in very-high-frequency and ultrahigh-frequency communications has resulted in the use of crystal control at relatively low frequencies followed by frequency multipliers and power amplifiers to obtain the desired output characteristic. Each intermediate stage can radiate on its fundamental as well as on the harmonics of the output frequency. In addition, circuit resonances at frequencies removed from the operating point may result in generation of parasitic oscillations. Improperly designed or tuned transmitters may produce spurious side bands when modulated.

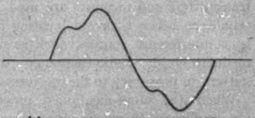
A microwave radar transmitter is seldom a problem to communications since its frequency is sufficiently removed to preclude interference to receivers other than those operating in the same band, and filters can readily be applied where overloading causes spurious responses in the receivers. On the other hand, the



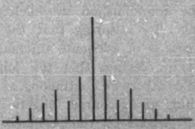
(A) RADAR WAVESHAPE
(BROADBAND)



(B) RADAR CARRIER OUTPUT



(C) COMMUNICATIONS WAVESHAPE
(CONTINUOUS WAVE)



(D) CW TRANSMITTER OUTPUT

Fig. 3. Typical interference wave shapes

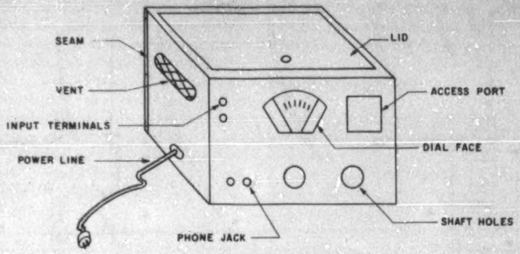


Fig. 4. Shielding enclosure leakage paths

modulator is a most prolific source of interference. The high peak powers and steep wave fronts produced by the pulsed output give rise to an extensive spectrum of spurious frequencies, as previously noted.

Not often suspected but a frequent offender is the superheterodyne receiver. Local-oscillator radiation from the antenna or chassis may be sufficient to block adjacent receivers on certain frequencies.

There are a number of other sources of interference, such as ignition systems, rotating machinery, and contacting devices, but a discussion of these factors cannot adequately be treated within the scope of this paper.

INTERFERENCE REDUCTION

Several preventive measures are available to the designer interested in obtaining equipment which is free of interference. In many cases interference-free components are available for assembly in complete equipments. Careful choice in component selection can do much to reduce the over-all suppression engineering. When installing systems where both transmitters and receivers are used, care should be taken to avoid locating high-level interference sources, such as radar modulators or radio-frequency transmitters, in proximity to the receivers.

When the afore-mentioned measures are insufficient, there are other means of controlling the interference. The primary methods by which such suppression may be accomplished are shielding and filtering. By ensuring that the equipment case is of adequate thickness, and by making certain that the joints are welded or otherwise continuous, as well as providing spring-contact fingers or conducting gaskets on the periphery of removable covers or doors, the induction and radiation fields can be contained within the equipment enclosure.

To maintain the shielding integrity, all lines entering or leaving the enclosure must be decoupled by suitable filters. These may consist of simple feed-through capacitors, or of more complex inductance-capacitance networks. In the presence of strong radiation at a fixed frequency, such as the microwave energy from a radar transmitter, tuned filters in the receiver antenna circuit are often quite effective. It is also necessary to make sure that all units are bonded to ground with a strap that offers low impedance to radio frequencies.

Equipment Design

It is well to consider some specific examples illustrating application of the foregoing techniques to equipment design.

RADAR MODULATORS

There are two principal sources of interference from a radar: the modulator and the radio-frequency transmitter. The radar modulator can produce considerable pulse energy at radio frequencies, which may cause severe interference to near-by communications equipment. In the design of the modulator there are three areas in which suppression must be accomplished:[1]

1. Enclosure design, to attenuate radiated interference.
2. Filter design, to reduce conducted interference.
3. Design of the high-voltage pulse cable and connectors.

An ideal enclosure should have no openings and should be of sufficient thickness to preclude case penetration by the interference energy. In practice, there must be openings to provide ventilation and permit accessibility for maintenance; see Fig. 4. In addition, there may be discontinuities in the shield for signal and power leads, control shafts, etc. To maintain the shielding effectiveness requires careful attention to all these possible sources of leakage.[2]

The following measures will help to maintain the desired shielding effectiveness:

1. Since seams which are riveted or bolted are ineffective for shielding at radio frequencies, they should be welded, brazed, or soldered such that the joint is continuous.
2. Removable panels or covers should have the finish removed from all mating surfaces. Spring-contact fingers or conductive gaskets will help provide good electrical contact around the periphery of the panel.
3. All openings for ventilation or drainage over 1/16 inch diameter should be screened with fine copper mesh. As an alternative, the opening may be designed with a sleeve which acts as a wave-guide attenuator when properly dimensioned.
4. Conductors which must carry pulse wave forms should be separated from low-level or d-c cables to prevent mutual coupling. The pulse leads should be well shielded and terminated in low-impedance connectors which are separate from all other leads. (One technique that has been adopted for an air-borne equipment was to mount the magnetron directly on the pulse-transformer housing, thus eliminating the need for a pulse cable between the modulator and transmitter.)

To prevent interference from being conducted out of the case on connecting leads, all leads entering or leaving the enclosure should be well filtered. Often the use of feed-through capacitors will give sufficient interference reduction; in severe instances, however, it will be necessary to use low-pass filters to decouple the noise voltages. To preclude coupling between input and output leads, they should be shielded and isolated so that interference will not be reintroduced on conductors leaving the equipment. It is also good practice to use well-grounded electrostatic shields between the high-voltage and low-voltage windings of power or filament transformers.

RADAR TRANSMITTERS

Because of the high peak power levels, a microwave radar transmitter can cause considerable interference to communications receivers, even if they are not tuned to the interfering frequency. This is caused by the Fourier components of the modulation frequency, which may produce side bands extending over hundreds of megacycles. Such interference may be coupled directly from the antenna or through case radiation. Case radiation may be considerably reduced by applying the shielding techniques previously described. The most effective measures for

Vol. 7
Pt. 1
1956

reducing antenna-radiated interference are to utilize wave forms containing only the minimum high-frequency harmonics actually required for the desired equipment performance and the reduction of receiver susceptibility. The latter problem will be considered in a later section.

COMMUNICATIONS TRANSMITTERS

While communications transmitters are not often susceptible to interference, they may produce considerable harmonic and spurious radiation. This is often caused by the need for getting the maximum power output from a given equipment within a limited space, which necessitates conserving the number of frequency-multiplying stages and, sometimes, overdriving the final amplifier.

An unstable oscillator is very susceptible to parasitic oscillations. Where practical, crystal control should be used and the oscillator should be loosely coupled to the following stage. Since capacitive coupling networks often permit the passage of harmonics and spurious radiations to succeeding amplifiers, a tuned coupling network is to be preferred. The power amplifier should not be used as a frequency multiplier, since it will contain a large undesired component at the fundamental frequency. The output stage should be designed for low distortion at the required power output to minimize the need for excessive grid excitation and heavy loading, both of which lead to the generation and transfer of spurious signals into the antenna.

If the foregoing measures have been followed, together with careful shielding and decoupling of the individual stages, the harmonic and spurious content of the radio-frequency carrier can be reduced to a low level. For further interference reduction, it may be necessary to insert a low-pass filter in the output circuit to reduce harmonic radiation The application of such design techniques will permit reducing undesired interference signals to as much as 80 decibels less than the carrier amplitude.

When installing the transmitter, care should be taken to avoid routing antennas and transmission lines near susceptible receivers. The use of shielded transmission lines will help considerably to reduce coupling from the antenna system.

RECEIVERS

While considerable interference reduction can be achieved at the source of the interference, much can be done to reduce receiver susceptibility. In some cases even properly designed and operated transmitters can produce interference or spurious signals in a receiver because of its nonlinearity or undesirable responses.

A well-designed receiver case should prevent the entrance of any signal except through the antenna. The case-shielding methods previously described for transmitters are also applicable to receivers. The use of shielding and decoupling of leads entering the receiver will help greatly to reduce receiver susceptibility caused by case penetration.

It is also important to reduce receiver bandwidth to the minimum required for reception of the desired signal. This will measurably assist in rejecting adjacent channel interference. Other design features which may reduce the interference or its effect on the output are:

1. High image-rejection ratio.
2. Balanced antenna input circuit.
3. Use of peak noise limiter.
4. Shielded antenna lead-in.
5. Tunable or fix-tuned antenna wave trap.
6. Receiver blanking for the duration of the interfering pulse.

The local oscillator of superheterodyne receivers may often be a source of interference to other receivers. Such interference can be reduced by operating the oscillator at minimum power level, by using internal shielding of the converter and oscillator stages, by employing electrostatically shielded coupling transformers, and by incorporating a radio-frequency amplifier stage.[3] The measures which can be applied successfully will, of course, depend on the operating characteristics of the system in which it is employed, e.g., radar, communications, navigation, etc.

Measurements

INSTRUMENTS

To permit evaluation of the degree of interference produced by an equipment and the effectiveness of interference-reduction measures, it has been necessary to develop standardized instruments and measuring techniques. For obtaining relative measurements it is common practice to use audio output meters which respond to the average or rms value of the receiver output. For engineering purposes, however, it has been found desirable to use an interference measuring set which responds to the peak value and provides information on the amplitude and spectral distribution of the interference. Such intruments, commonly called noise meters, usually consist of an antenna, amplifier, detector, indicator, and means for calibration.

The amplifier is a superheterodyne receiver preceded by an attenuator. Design requirements for such amplifiers are high sensitivity and signal-to-noise ratio, stability, constant bandwidth, accurate attenuators, and wide dynamic range. The choice of bandwidth depends on the type of interference signal under measurement; in general, the equipment must be capable of measuring both broad-band and narrow-band interference. Commercial instruments are now available which enable measurements to be made from 15 kc to 1,000 megacycles.

For measuring the peak value of broad-band interference, a slide-back vacuum-tube voltmeter is generally used. Some instruments have provision for both aural and visual null slide back. For continuous-wave measurements an average reading meter is provided. While a number of meters are also capable of reading "quasi peak," the difficulty of interpreting the reading has led the military services to adopt peak as the standard.

A calibrator is used to provide a standard voltage source with which the interfering signal may be compared. The unit of measurement is microvolts per unit bandwidth, which makes an allowance for receiver bandwidth variations. Much broad-band interference is of the impulse type and has been found to produce a response in the receiver proportional to the bandwidth. For such measurements an impulse noise generator has been developed which provides useful output as high as 1,000 megacycles. Some instruments utilize random noise calibration. The noise source is usually a shot diode or a gas tube. Calibration for continuous-wave measurements is made with a sine-wave generator operating at the input frequency of the amplifier.

Both vertical rod and horizontal dipole antennas are used when making tests of radiated noise, depending on the frequency of measurement. For locating point sources of interference, such as leakage in a cabinet, loop antennas are quite useful; however, difficulty in correlating their output with that of rod antennas precludes their use in field-intensity measurements.

For conducted interference the instrument is used as a 2-terminal radio-frequency voltmeter. A line-stabilizing network is provided to standardize the impedance of the line across which the interference voltage is measured.

Figs. 5 and 6 illustrate the placement and connection of the noise meter when used to make conducted and radiated interference measurements respectively.

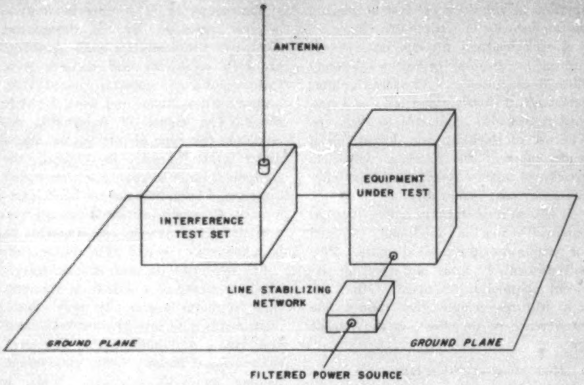


Fig. 5. Test arrangement for conducted interference measurements

TEST AREAS

For making interference measurements where the ambient interference level may be quite high, as in industrial areas, a suitable screen room must be used. Such rooms are available in a number of sizes and several types of construction, depending on the degree of attenuation required. The walls may be constructed of copper mesh, solid copper or iron, or a composite of these materials. The simplest structure consists of a single-wall shielding enclosure; where higher ambient levels are encountered, a double-shield construction is employed. This is most commonly of the cell type with panels of the enclosure shielded individually and the cells then joined together. Another method of construction utilizes two walls completely insulated from each other except at one common ground point. To maintain the shielding effectiveness of the screen room, the electric power must be brought through a line filter which will reject interference on the power line.

In many cases, reflections from a screen room may prove quite troublesome in obtaining useful readings. In addition, some equipments may be too large to measure in a screen room of reasonable size. In such instances, it is necessary to choose a suitable site in an open area where the ambient interference level is relatively low over the frequency range of interest. Some work is currently being conducted toward correlating screen room measurements with those made in open areas.[4]

Specifications

To insure uniformity in limits and methods of measurement for electronic equipment procured by the military services, a number of radio-interference specifications have been issued, covering specialized applications of electronic equipment.

Specification *MIL-S-11748 (Sig C)*[5] is employed by the U. S. Army for testing electronic equipment. Radiated measurements are required from 0.15 to 1,000 megacycles at test distances based on the minimum expected distance from susceptible communication equipment. Conducted measurements are made over the range of 0.15 to 40 megacycles.

MIL-I-6181B[6] is employed by the U. S. Air Force and Navy Bureau of Aeronautics for interference tests on electric and electronic equipment. A number of objective tests are required to evaluate the extent of generation or susceptibility of an equipment to interference. Radiated measurements are required from 0.15 to 150 megacycles for equipment capable of producing broad-band interference. The Air Forces have found that suppression over this range will generally provide satisfactory interference reduction up to 1,000 megacycles.

On the other equipments, including those emitting continuous-wave and pulsed-continuous-wave radiation, measurements are required from 0.15 to 1,000 megacycles. The antenna is located 1 foot from the equipment. Conducted interference measurements are made from 0.15 to 20 megacycles. Measurements are not required on leads interconnecting the various units of a given system.

On some equipments, tests are also required for susceptibility to conducted audio interference and to conducted and radiated radio-frequency interference. Finally, measurements are made of spurious front-end receiver responses and spurious antenna radiation from transmitters and receivers.

For ship-borne or ground-based electronic equipment, *MIL-I-16910A (Ships)*[7] is utilized by the Navy Bureau of Ships. This specification covers radiated measurements over the frequency range of 14 kc to 1,000 megacycles, and conducted interference from 0.15 to 100 megacycles. The antenna is located 3 feet from the equipment under test for measuring radiated interference. Specifications are provided for receiver oscillator radiation and spurious transmitter radiation.

Conclusions

It has been attempted to present the major interference problems confronting the designer of electronic equipment, and remedial measures have been suggested. In brief, the latter include:

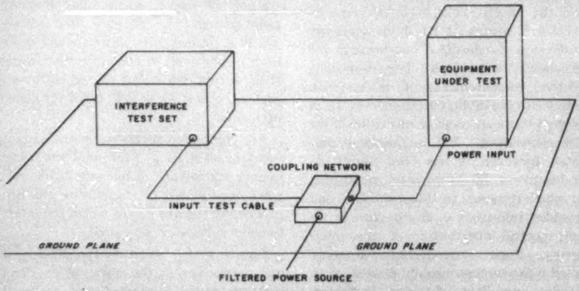


Fig. 6. Test arrangement for radiated interference measurements

1. Case shielding.
2. Filtering of input and output leads.
3. Good radio-frequency bonds.
4. Conservative design.

The importance of considering interference reduction in the initial design cannot be stressed too strongly. Although interference suppression generally goes hand in hand with good design practices, it must also be considered in its own light at a very early stage to be most effective.

References

1. An Investigation of Interference from Radar Modulators, J. P. Quine, H. Q. Totten. *Contract AF 30(602)-401*, Rensselaer Polytechnic Institute, Troy, N. Y., 1954.

2. Problems in Shielding Electronic Equipment, C. S. Vasaka. *Proceedings*, Conference on Radio Interference Reduction, Chicago, Illinois, Dec. 1954, pp. 86–103.

3. Radio Interference Suppression Techniques. *Contract DA-36-039sc-42596*, Coles Signal Laboratory, Fort Monmouth, N. J., pp 139–141.

4. An Investigation of Problems Associated with Broad Band Interference Measurements. K. E. Mortenson, C. J. Truex. *Contract No. 36-039 sc-42488*, Rensselaer Polytechnic Institute, Troy, N. Y., 1954.

5. Suppression Requirements for Electrical and Electronic Equipment (Except Internal Combustion Engine-Driven Equipment). *MIL-S-11748 (Sig C)*, U. S. Army, Washington, D. C., Feb. 14, 1952.

6. Interference Limits, Test and Design Requirements. Aircraft Electrical and Electronic Equipment. *MIL-I-6181B*, U. S. Air Force, Washington, D. C., May 29, 1953.

7. Interference Measurement, Radio, Methods and Limits; 14 Kilocycles to 1000 Megacycles. *MIL-I-16910A (Ships)*, U. S. Navy Bureau of Ships, Washington, D. C., Aug. 30, 1954.

Method of Measuring Spot Size of Cathode-Ray Tube

R. B. KUHN
NONMEMBER AIEE

D. LEVINE
NONMEMBER AIEE

THE determination of the spot size of a cathode-ray tube has been, at best, a subjective process. Thus, the shrinking-raster method of measuring spot size relies on the judgment of the observer to determine when a set of lines appears to merge. Similarly, any attempt to measure the spot diameter with a microscope is also dependent upon the observer because of the way the flux density varies over the area of the spot.[1]

This paper describes an objective method for determining spot size. The spot size is defined as being equal to the side of a square within which the current density in the Gaussian distribution about a stationary electron beam $I_c\epsilon^{-k^2r^2}$ may be redistributed uniformly with current density I_c.[2] Given this definition, the constant k is expressed as

$$k=\frac{\sqrt{\pi}}{w}=\frac{1.7725}{w} \qquad (1)$$

Nomenclature

$H(x)\equiv\frac{2}{\sqrt{\pi}}\int_0^x \epsilon^{-u^2}du$ = probability integral
I_c = current density at center of electron beam
$I_c\epsilon^{-k^2r^2}$ = current density anywhere in electron beam
k = constant determined by definition of spot size for Gaussian spot; it is equal to $\sqrt{\pi}/w$ herein
L_0 = reference level of light flux, measured when flying spot is focused on center of widest or reference slit on test mask
L_M = light flux incident upon photocathode when flying spot is focused on center of slit other than the reference slit
L = light flux incident on photocathode after passing through slit
$r^2=x^2+y^2$
W = width of largest slit and each opaque dividing strip
w = width of the cathode-ray-tube spot
x = distance from center of spot to leading edge of slit
y = direction of long dimension of a slit
λ = width of any slit
ρ = light flux density at center of spot
$\rho\epsilon^{-k^2r^2}$ = light flux density anywhere in the spot

Instrumentation

The method requires an instrument, Fig. 1, consisting chiefly of a mask, an objective lens to focus the cathode-ray-tube spot onto the mask, and a condensing lens to focus the light passing through the mask onto a phototube. The object of the condensing lens is the exit pupil of the objective lens. A filter is placed in the light beam to pass only the short persistent component of the phosphor light radiation. In practice, the filter would undoubtedly be mounted on the object lens; however, its position is unimportant. The instrument may be designed so that the measured value of the spot size has a rms uncertainty of less than 4 per cent.

The mask consists of an opaque background broken by slits of various dimensions; see Fig. 2. The opaque area between any two slits, as well as the largest slit, each has a width W at least four times the width of the largest spot to be measured. The width of the narrowest slit is approximately four-fifths the size of the smallest spot to be measured. The length of each slit is 20 times the width of the largest spot to be measured. At the time that this paper was prepared, a prototype instrument similar to the one shown in Fig. 1 had been built, but no actual measurements had been made.

Theory

When the spot is focused within a single slit of width λ, the relative light output, neglecting phosphor decay, is given by

$$\frac{L}{L_0}=\frac{1}{2}\{H[k(x+\lambda)]-H(kx)\} \qquad (2)$$

See the Appendix for derivation of equation 2.

Equation 2 has its maximum value when

$$x=-\frac{\lambda}{2} \qquad (3)$$

i.e., when the spot is centered in the slit. If the maximum is denoted by subscript M, then

$$\frac{L_M}{L_0}=\frac{1}{2}\left[H\left(\frac{k\lambda}{2}\right)-H\left(-\frac{k\lambda}{2}\right)\right]=H\left(\frac{k\lambda}{2}\right) \qquad (4)$$

Substitution of the definition of k into equation 4 yields

$$\frac{L_M}{L_0}=H\left(\frac{\sqrt{\pi}}{2}\,\frac{\lambda}{w}\right) \qquad (5)$$

The light passing through a slit is measured by placing the output of the phototube on a suitable recorder. The light through a mask, Fig. 3(A), as

---

Paper 56-265, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted June 16, 1955; made available for printing December 20, 1955.

R. B. Kuhn and D. Levine are with the Goodyear Aircraft Corporation, Akron, Ohio.

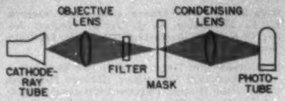


Fig. 1. Representation of spot-size measuring instrument

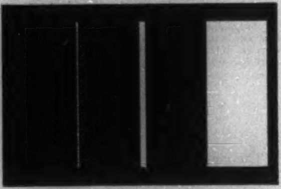


Fig. 2. Mask for spot-size measurement. Length of each slit is 3.588 inches; widths, from left to right, are 0.007, 0.021, and 0.063 inch. Dimensions are for mask used with prototype instrument

recorded at the phototube output is illustrated in Fig. 3(B) for the case where the spot size is equal to the width of the second slit. When the spot is in the center of the first slit, there is no significant limiting of the light output, and the value of L_0 is determined; see Fig. 3(C), left. However, when the spot is at the center of a slit that is equal to the spot size, 21 per cent of the light is obstructed by the opaque area; see Fig. 3(C), center. In Fig. 3(C), right, the light contours permit the further reduction of maximum light associated with a slit of half the spot size to be visualized.

After L_M/L_0 is determined, equation 5 may be used to solve for spot size. Alternatively, the graph of Fig. 4 may be employed. The curve drawn here has a point of inflection at

$$\frac{w}{\lambda}=\frac{\sqrt{\pi}}{2}=0.88623 \tag{6}$$

as derived elsewhere.[3] Consequently, in the region of this spot-size value, a small error in the determination of the light ratio has a minimum effect on the accuracy of the spot-size measurement.

Error Analysis

The maximum relative error is given by

$$\frac{\Delta w}{w}=\left(\frac{w}{\lambda}\right)\epsilon^{\frac{\pi}{4}\left(\frac{w}{\lambda}\right)^{-2}}\left(\frac{\Delta L_M}{L_0}-\frac{L_M}{L_0}\,\frac{\Delta L_0}{L_0}\right)+\frac{\Delta\lambda}{\lambda} \tag{7}$$

as shown elsewhere.[3]

For a 1-per-cent error in slit width and in each recorded reading, the maximum relative error and rms error are expressed as

$$\frac{|\Delta w|}{w}(\max)\leqq 0.02\,\frac{w}{\lambda}\,\epsilon^{\frac{\pi}{4}\left(\frac{w}{\lambda}\right)^{-2}}+0.01 \tag{8}$$

and

$$\frac{\Delta w}{w}(\mathrm{rms})\leqq\left[0.0002\left(\frac{w}{\lambda}\right)^2\epsilon^{\frac{\pi}{2}\left(\frac{w}{\lambda}\right)^{-2}}+0.0001\right]^{1/2} \tag{9}$$

as derived elsewhere.[3]

The error functions are plotted in Figs. 5 and 6 for either 1- or 2-per-cent maximum error in each reading. From the graphs it follows that the minimum relative error in the measured spot size is obtained when the ratio of the spot size to the slit width is 1.25; for a 1-per-cent recording and reading error, the rms relative error is less than 4 per cent over the range

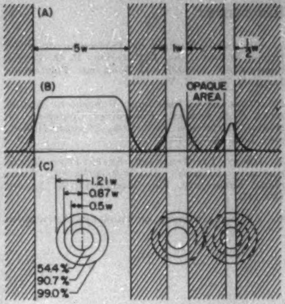


Fig. 3. Test chart function

A—Test mask
B—Light variation for swept beam
C—Light distribution about centered spot

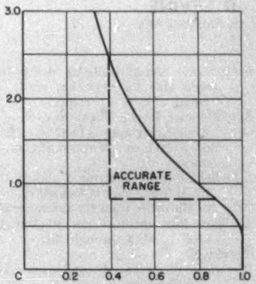


Fig. 4. Spot-to-slit ratio w/λ as function of light ratio L_M/L_0

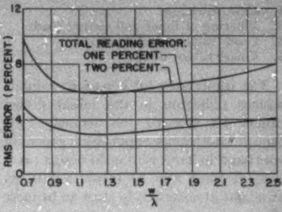


Fig. 5. Rms value of relative spot-size error as function of spot-to-slit ratio w/λ

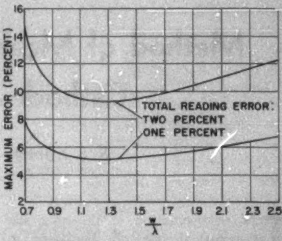


Fig. 6. Maximum value of relative spot-size error as function of spot-to-slit ratio w/λ

$$0.8<\frac{w}{\lambda}<2.4 \tag{10}$$

The maximum error is 6.5 per cent over this range, which corresponds to the accurate range region specified in Fig. 4. Consequently, a mask designed for measurement of a wide range of spot sizes may have adjacent slit widths in a 3-to-1 ratio.

It also follows from Figs. 5 and 6 that, for a 2-per-cent recording and reading error, the rms relative error is less than 8 per cent over approximately the same range as equation 10, while the maximum relative error is 12 per cent.

Since the term $\Delta\lambda/\lambda$ in equation 7 is constant, the errors between relative readings with the same instrument are dependent mainly on the recording apparatus. Thus, when one instrument is used for all determinations, relative spot sizes can be measured more accurately.

The basic relation of this report is equation 2, which was derived on the assumption that the distribution of the electrons in a beam is Gaussian, and that the light output of the phosphor (neglecting decay) is directly proportional to the current density. Consequently, the error equations will apply only when these assumptions are closely approximated.

Conclusions

If the light flux density of the cathode-ray tube has a Gaussian distribution, the spot size, as defined in the foregoing, can be determined in an objective manner. The accuracy of the determination is dependent on the ratio of the spot size to the slit width, errors in the recording apparatus, and errors in the slit-width measurement.

Appendix. Derivation of Relative Light Output

When the cathode-ray-tube spot is focused on a slit of width λ, the light passing through the slit is expressed as

$$L=\rho\int_{x}^{x+\lambda}\epsilon^{-k^2x^2}dx\int_{y}^{y+5W}\epsilon^{-k^2y^2}dy \qquad (11)$$

The integrals of equation 11 are converted to normal probability integrals by the substitutions

$$y=\frac{w}{k}$$

and

$$x=\frac{v}{k}$$

The result is

$$L=\frac{\pi}{4k^2}\{H[k(x+\lambda)]-H(kx)\}\times \{H[k(y+5W)]-H(ky)\} \qquad (12)$$

Since $5W$ is 20 times the width of the largest spot to be measured, the second factor of equation 12 is at least as large as

$$[H(ky+20\sqrt{\pi})-H(ky)]$$

For all practical purposes, the latter term is equal to 2 when the spot is nearly centered in the y direction. Thus equation 12 becomes

$$L=\frac{\pi}{2k^2}\{H[k(x+\lambda)]-H(ky)\} \qquad (13)$$

When λ is four times the width of the largest spot to be measured, the amount of light passing through the largest slit is given by

$$L_0=\frac{\pi}{2k^2}[H(kx+4\sqrt{\pi})-H(kx)] \qquad (14)$$

With the spot centered in the x direction, the quantity in brackets is, for all practical purposes, equal to 2. Thus, equation 14 becomes

$$L_0=\frac{\pi}{k^2} \qquad (15)$$

All the slits are $5W$ long. Therefore, the light passing through any single slit is given by equation 15, and the ratio of L to L_0 is expressed as

$$\frac{L}{L_0}=\frac{1}{2}\{H[k(x+\lambda)]-H(kx)\} \qquad (16)$$

References

1. Cathode-Ray Tube Displays, T. Soller, M. A. Starr, G. E. Valley, Jr. McGraw-Hill Book Company, Inc., New York, N. Y., 1948, pp. 590–600.

2. Significance of Line Width in PPI Displays, D. Levine. *GER-6142*, Goodyear Aircraft Corporation, Akron, Ohio, June 16, 1954, p. 37.

3. An Instrument for Measuring Spot Size, R. B. Kuhn, D. Levine. *GER-5670, Ibid.*, Jul. 26, 1954, equations A-14, p. 16; A-24, p. 19; A-27, p. 20.

◆

Discussion

G. R. Spencer (Philco Corporation, Lansdale, Pa.): The method described in this paper provides an answer which is meaningful only to the extent that the light flux density distribution approximates the Gaussian law, but the authors do not reveal here any means for determining how nearly their distributions approach Gaussian curves. Indeed there are several factors which tend to cause departures from the true Gaussian distribution, the most important of which are electron lens aberrations, space charge conditions, aperture masking, and phosphor saturation. Further, misalignment of gun parts and stray magnetic fields frequently cause marked lack of circular symmetry.

It seems appropriate, therefore, to point out that an instrument which provides a picture of the brightness distribution across a line traced by the electron spot would be of far greater value in most applications. Such an instrument has been described adequately in the literature,[1] and some equipments using this principle have been used at Philco, and perhaps elsewhere, for several years. Fig. 7 illustrates a typical output signal in which the brightness distribution across two adjacent lines of one field of a television raster is represented. The known line spacing provides a convenient calibration datum.

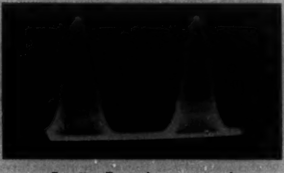


Fig. 7. Typical output signal

It is conceivable that the answers obtained with the apparatus described by the authors, in spite of the uncertainties as to their precise meaning, may be useful in some applications. Will the authors suggest applications for which their method would give meaningful results? And for what ranges of beam currents and anode voltages has this technique been found practicable?

Reference

1. Evaluating Performance of TV Picture Tubes, Julius Green. *Electronics*, New York, N. Y., vol. 25, no. 2, Feb. 1952, pp. 124–29.

R. B. Kuhn and D. Levine: We appreciate Mr. Spencer's comments and his calling attention to Mr. Green's apparatus.[1] The introduction to our paper would have been written differently had this reference been known to us earlier.

In view of the well-known variety of spot shapes that are to be found under certain conditions,[2] it was decided to concentrate on properly defined spots in the design of our equipment. It should be realized that this viewpoint taken represents that of a "using" organization as contrasted with one that is concerned with tube development. Consequently, if a spot is seriously deformed because of poor electron optics, it would result in rejection of the tube (on the part of the user) without the necessity of taking measurements.

While the analysis of our instrument is limited to a Gaussian variation of light flux, alternative derivations can be provided for any other known distribution. In effect, though, one must start with information concerning the flux distribution, and the assumption made appeared the most reasonable following a literature survey.[3] Consequently, our equipment admittedly finds no application to studies of the influence of various factors on the shape of the spot. For this purpose the instrument developed by the Philco Laboratories appears to be ideal.

Basically, both developments have as their aim the providing of an objective measure of spot size to replace the subjective shrinking-raster method or other visual means. To achieve this goal, both approaches use aperture techniques. Mr. Green has designed his apparatus about an aperture that is very narrow relative to the spot size, while our equipment employs an aperture having a width of the same order of magnitude as the spot size; see equation 10.

The first-mentioned approach has the advantage of providing detailed information about spot shape, while the latter technique requires very uncritical sweep positioning. This factor is of importance to our immediate application, which is to check cathode-ray tubes installed in radar plan-position indicators.

References

1. See reference 1 of Mr. Spencer's discussion.

2. The Image Formation in Cathode-Ray Tubes, G. Liebmann. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 34, Aug. 1946, pp. 580–86.

3. See reference 1 of the paper, especially Fig. 3.4, p. 97, and Fig. 3.16, p. 105.

Two New Total Radiation Pyrometers

W. DERGANC
ASSOCIATE MEMBER AIEE

S. N. HOWELL
NONMEMBER AIEE

Synopsis: The measurement and control of temperature are basic to all industrial processes. Modern automatic machinery exploits the economy achieved by a continuous flow of product. Such machinery requires instrumentation capable of measuring the temperature of moving material in process for maximum efficiency. Total radiation pyrometers are useful for this purpose, and there is a continuing demand from industry for more sensitive, accurate, and versatile instruments. This paper describes two new pyrometers* applicable to a wide variety of industrial and laboratory measurements. They combine good spectral response, high sensitivity, and rapid response, with stability and accuracy adequate for most requirements. These factors are discussed in relation to the design of the pyrometers. The basis for transferring a radiation measurement to a temperature by calibration and some of the limitations involved are also considered.

Brief Description of Instruments

TYPE IR-2 SYSTEM

THE first instrument to be described is the Servotherm† type *IR-2* radiation pyrometer system, shown in Fig. 1, and consisting of a pyrometer head, an amplifier, and power supply. This is a general-purpose unit for measurements in the range from 60 to 1,300 C (degrees centigrade) and is capable of maintaining calibration under the ambient conditions encountered in manufacturing plants. The output is directly applicable to proportional automatic control systems. Fig. 2 is a simplified section of the pyrometer head, and Fig. 3 shows a partial schematic of the amplifier and power supply circuits. The system functions as follows.

Incoming radiant energy is collected by the Servofrax (arsenic trisulfide glass) lens and, after interruption by the mechanical chopper at a 15 cps (cycles per second) rate, is focused on a Servotherm thermistor bolometer detector. This energy is compared with similarly interrupted energy from an internal reference radiation source via a second detector, bridge-connected with the first. The bolometer bridge is biased with direct voltages balanced to ground, and the cyclic interruption of the energy produces cyclic variations of the bolometer resistances. Any difference in energy levels between incoming and reference radiations produces an alternating voltage output from the bridge to the preamplifier grid. The phase of this output voltage depends on which radiation level is higher, since the radiation from either chopper itself is equal to the ambient level of the pyrometer head.

After amplification by the preamplifier and main amplifier units, the output is rectified by a synchronous rectifier on the chopper shaft. This produces a direct voltage polarity sensitive to the energy levels; i.e., a source hotter than the reference produces a positive voltage output, and a colder one, a negative.

The internal reference source is always used to null the incoming radiation, making it possible to utilize full amplification for high-differential sensitivity at all times. Once the pyrometer is set for a given temperature, small deviations from it appear as relatively large voltage outputs which may be used for manual indication and control or as the input signal to an automatic system.

The field of view of this system is one degree square and the lens focuses from 1.5 feet to infinity. The system time constant, once set at a given temperature, is approximately 250 milliseconds. The time required for the reference source to stabilize at a new setting varies from about 30 seconds to five minutes depending upon the initial and final conditions.

TYPE IR-1 SYSTEM

The type *IR-1* system is a fast and very sensitive general-purpose laboratory instrument for measurements from ambient to 1,000 C. It employs essentially the same electronic system as the *IR-2* but a different pyrometer head; see Figs. 4 and 5. In this system incoming energy is collected by a first-surfaced spherical mirror. The energy is interrupted by a motor-driven chopper at 180 cps and focused on a thermistor bolometer detector. This detector is bridge-connected with a compensating thermistor in the same housing and its output is amplified and rectified as for the *IR-2* system. The output is proportional to the difference in radiant energy between the chopper radiating at ambient temperature and the incoming radiation. Focusing is accomplished by moving the detector-preamplifier assembly, and a scale covering from two feet to infinity is provided. The system covers a 1-degree square field, and the over-all response time is 25 milliseconds. A feature of this unit is the parallax-free periscope sight which permits accurate aim in detecting hot spots, etc. The sight contains a built-in output meter so that simultaneous viewing of target and output signal is possible. Illumination balance is achieved by a controlled meter light and adjustable polaroid filter disks for the target.

Design Considerations

SPECTRAL RESPONSE

Available Energy

Fig. 6 shows the relative energy versus wave length for blackbodies at temperatures between 20 and 920 C. Note that 75% (per cent) of the total energy, represented by the area under the curves, is on the long wave-length side of the radiation maximum, and that most of the total energy is at wave lengths greater than one micron. For example, at 1,000 C, less than 0.5% of the total energy is below that wave length. At lower temperatures, considerably less total energy is available, and the distribution shifts to longer wave lengths. To utilize most of the available energy, a pyrometer should detect the energy from 1 to 12 microns or more. In some cases, such as the determination of total emissivity, a response to 25 microns is required for accurate results.

The spectral sensitivity of the pyrometers is limited by three factors: the collecting optics, detector window, and detector response. The last-mentioned is relatively flat in the 1- to 25-micron region so that the first two items are the essential limitations in this case.

IR-2 Optics

For the *IR-2* system, arsenic trisulfide glass was selected for the lens. As

---

* Products of Servo Corporation of America; patents applied for.

† Registered trade-mark, Servo Corporation of America.

Paper 56-98, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted August 5, 1955; made available for printing March 15, 1956.

W. DERGANC and S. N. HOWELL are with the Servo Corporation of America, New Hyde Park, N. Y.

Vol.
Pt. 1
1956

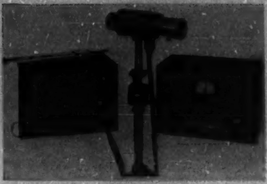


Fig. 1. Type IR-2 pyrometer system showing power supply (left), pyrometer pickup head (center), and amplifier unit (right)

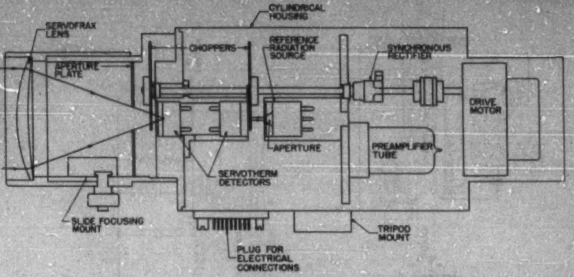


Fig. 2. Simplified sectional view of type IR-2 pyrometer head

shown in Fig. 7, this glass has good transmission in the 1- to 12-micron region. It is the first available material with far infrared transmission that can be fabricated in any reasonable size, that does not require special mounting or handling, and that is impervious to most chemicals. It has other desirable properties of a true glass, such as ease of cleansing.

The lens is a meniscus type, two inches in diameter, and *f 1.25*. The design was adjusted to keep the on-axis aberrations under the desired 1-degree field of view. Larger fields, up to about five degrees may readily be achieved by increasing the detector dimensions. Smaller fields may be accomplished by using a slower lens with reduced aberrations; this, however, reduces the over-all system sensitivity which, for given parameters, is inversely proportional to the *f* number of the lens. The detector window will modify the over-all response of this system. When arsenic trisulfide is used, the effect is to lower the over-all transmission while maintaining the same general shape of the curve. It should be pointed out that the arsenic trisulfide curve is for uncoated material. Coating can materially improve the transmission in the passband, but it is relatively expensive and has not been found necessary.

IR-1 Optics

This system utilizes an *f 0.8* first-surfaced collecting mirror of 3.5 inches focal length. The surface material is rhodium which has relatively flat and high transmission out to 25 microns so that the spectral response is limited by the detector window alone. Two typical response curves for such windows (other than arsenic trisulfide) are shown in Fig. 7. That for the coated silver chloride includes from four to 16 microns.

The relation of relative energy to temperature in pyrometer and blackbody systems is shown in Fig. 8.

DETECTORS

The detectors used in both pyrometers are of the flake thermistor type. They are housed in hermetically sealed capsules and they combine time constants of two milliseconds, excellent sensitivity, and uniform spectral response with ruggedness and long-term stability. The voltage output for square-wave interruption of energy is

$$V/\Delta W=K\left[1-\exp\frac{-t}{T}\right] \qquad (1)$$

where

V = voltage output for
ΔW = incremental radiation
T = detector time constant
t = time
K = A constant

For a typical detector as used in the *IR-2* system of 1 by 1 by 0.01 millimeter thick, the sensitivity is 250 volts per watt of incident radiation. The 15 cps interruption rate used in this system permits utilization of practically all of the available sensitivity, while in the *IR-1*, the response was compromised somewhat in favor of increased speed of response. When properly biased, the bolometer bridge may be operated within six decibels of theoretical noise which is the thermal noise of the thermistors. The rms noise output under typical conditions is about two microvolts at an impedance level of one megohm. A detector used in the *IR-2* system is shown in Fig. 9.

ELECTRONIC SYSTEM

Full exploitation of the available signal-to-noise ratio from the detectors requires extreme care in the design and construction of the electronic system. The detector bias must be noise free at the microvolt level, and insulation resistance in excess of 50,000 megohms to ground must exist at the input circuit to avoid extraneous noise currents. Amplification in the low-frequency region without introducing noise from a variety of sources calls for extremely well-filtered grid and plate voltages, minimum grid current, d-c filament supply, nonmicrophonic construction and components, and careful attention to grounding and shielding. The last-mentioned is particularly important in view of the proximity of the drive motors to the input circuit.

The preamplifiers built into the pyrometer heads meet these requirements. Low-noise, nonmicrophonic tubes and components are employed, and the operating conditions carefully adjusted for the best signal-to-noise ratio. Sufficient amplification is available to allow the use of a flexible shielded cable to the main amplifier unit.

The design of the main amplifier, Fig. 3, is conventional. Calibrated step attenuators are used to permit positive reset to a calibration point. The 3-stage feedback amplifier with a gain of about 7,500, drives a phase splitter whose output is fed to the synchronous rectifier. The d-c output from the latter is *RC*-filtered and fed to a differential output stage. This combination was found superior to the use of a selective amplifier tuned to the chop frequency. It includes energy components in the detector output, equation 1, over the fundamental, and eliminates the stability problem of synchronizing motor speed to amplifier tuning. Since a negative supply is used for the cathode returns, the output is both balanced and balanced to ground. Unbalanced output may be taken from terminal *A* to ground. The output stage will deliver about four milliamperes into 1,000 ohms and is designed to saturate on overloads to protect the recorder. Two special power supplies were designed to meet the re-

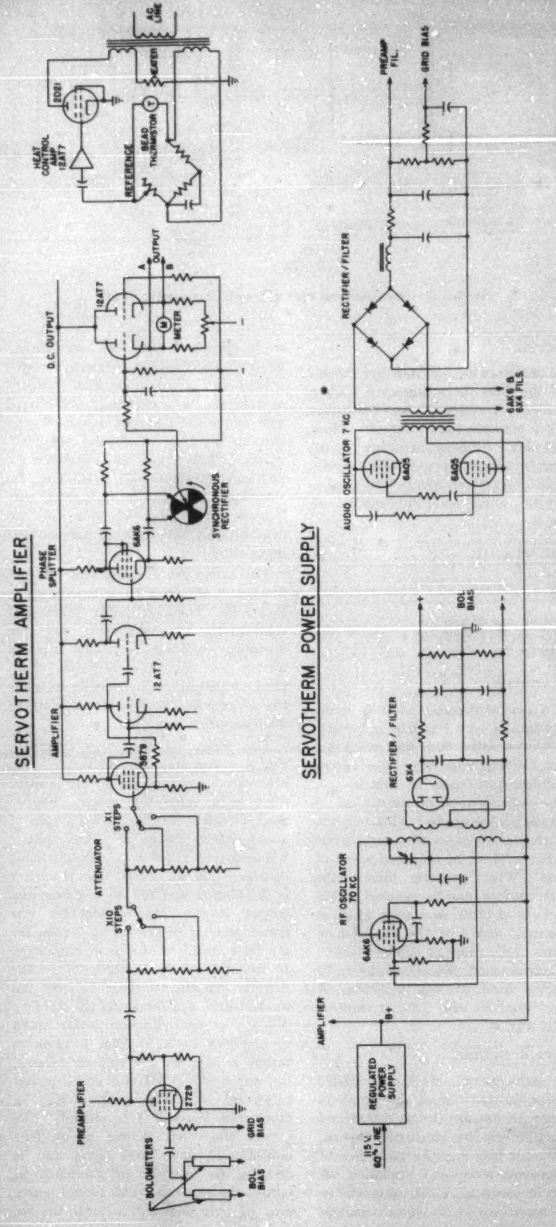

SERVOTHERM AMPLIFIER
PREAMPLIFIER
BOLOMETERS
BOL. BIAS
GRID BIAS
2729
X10 STEPS
ATTENUATOR
X1 STEPS
5879
AMPLIFIER
12AT7
PHASE SPLITTER
6AK6
SYNCHRONOUS RECTIFIER
D.C. OUTPUT
12AT7
METER
OUTPUT
A
B
HEAT CONTROL AMP 12AT7
2D21
REFERENCE
BEAD THERMISTOR
HEATER
AC LINE
SERVOTHERM POWER SUPPLY
115 V. 60~ LINE
REGULATED POWER SUPPLY
AMPLIFIER
B+
RF OSCILLATOR 70 KC
6AK6
RECTIFIER / FILTER
6X4
BOL BIAS
AUDIO OSCILLATOR 7 KC
6AQ5
6AQ5
6AK6 & 6X4 FILS
RECTIFIER / FILTER
PREAMP FIL.
GRID BIAS

Fig. 3. Partial schematics of amplifier and power supply units

quirements of well-filtered bolometer bias and preamplifier voltages. The bolometer bias supply uses a radio-frequency oscillator operating at about 70 kc. The oscillator output is fed to a carefully balanced rectifier-filter circuit. Since the plate supply to the oscillator is well regulated, output disturbances are minimized. Small residual variations appear balanced across the output. Since the bolometer bridge is balanced, these tend to cancel out. Filament power for this supply is also derived from a regulated source. These precautions plus careful attention to reducing 60-cycle pickup and microphonics make it possible to operate the biased bolometer bridge within six decibels of theoretical resistance noise for normal main supply variations from 105 to 125 volts.

The regulated filament supply uses a power audio-oscillator operating at 7 kc. Plate power for this unit is also well regulated. Part of the a-c output is used to power the bolometer bias oscillator and rectifier. The remainder is rectified and filtered and supplies d-c filament and grid bias voltages for the preamplifier.

Reference Radiation Source

The reference source, Fig. 9, as used in the *IR-2* system, utilizes a small nichrome-wound heater element and an arsenic trisulfide window. The radiation temperature of the heater is sensed by a small bead thermistor which is bridge-connected with the control potentiometer. The bridge is excited at line frequency and its output after amplification is fed to a thyratron which controls the power input to the source. The amplifier phase shift is arranged to produce proportional output control at the balance point. Extremely stable control of the source is obtained in this manner.

Two alternative sources have been employed. The first is similar to that just described but it utilizes a platinum-wire winding instead of the bead thermistor as the temperature sensing element. This eliminates the limitation on operating temperature imposed by the thermistor rating. The second is a combination heater and liquid-cooled unit which allows operation down to and below ambient temperature.

Calibration and Application

General

The use of total radiation pyrometers to measure temperature is, in theory, subject to many errors, in fact, to all of

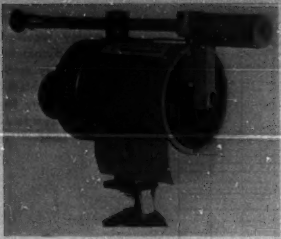


Fig. 4. Type IR-1 pyrometer head

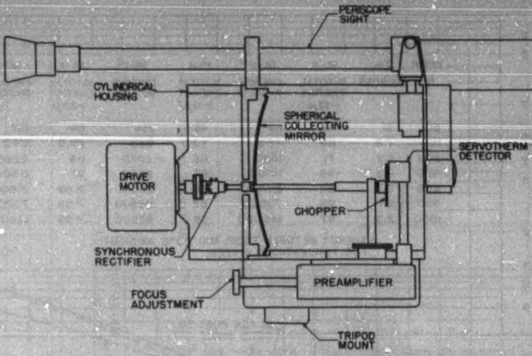


Fig. 5. Simplified sectional view of type IR-1 pyrometer head

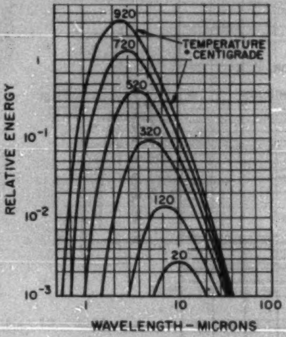


Fig. 6. Energy distribution of blackbody radiant energy with respect to wave length at different temperatures

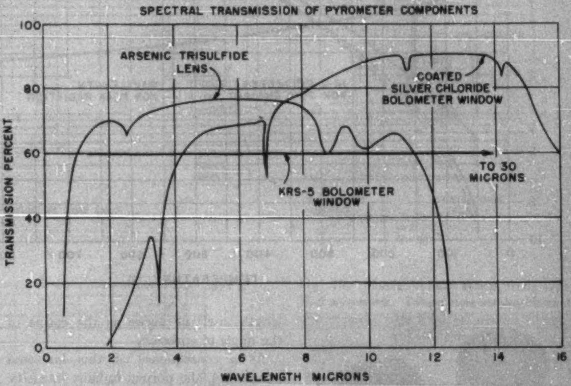


Fig. 7. Spectral transmission of pyrometer components

those found in measuring radiant energy, which suggests a formidable list of factors affecting the accuracy of such devices. In practice this is not the case. Simple calibration of the instrument under controlled conditions is usually adequate to produce accuracies comparable to other temperature-measuring devices used in the plant. A typical example is the manufacture of sheet film. This process required careful control of the temperature of film moving at high speeds. Thermocouples or thermistors cannot accurately show the film temperature. However, a stationary piece of film may be set up under controlled temperature conditions, and the pyrometer calibrated against this piece. This calibration may then be transferred to the film in process.

Such a transfer calibration is usually preferable to attempting a relationship to a standard blackbody calibration. The latter, however, is useful when emissivity factors are to be determined, and under other special conditions. A working knowledge of blackbody radiation is also required for the intelligent use of any total-radiation pyrometer, and a brief review of blackbody radiation laws follows.

BLACKBODY RADIATION

There are three bases for relating total radiation to temperature: the Planck equation defining the spectral distribution of energy, the Wien displacement law, and the Stefan-Boltzmann law of total radiation. These are expressed mathematically as follows.

$$\text{(Planck)}\quad W\lambda = C_1 \times \lambda^{-5}\left[\exp\frac{C_2}{\lambda T} - 1\right]^{-1} \tag{2}$$

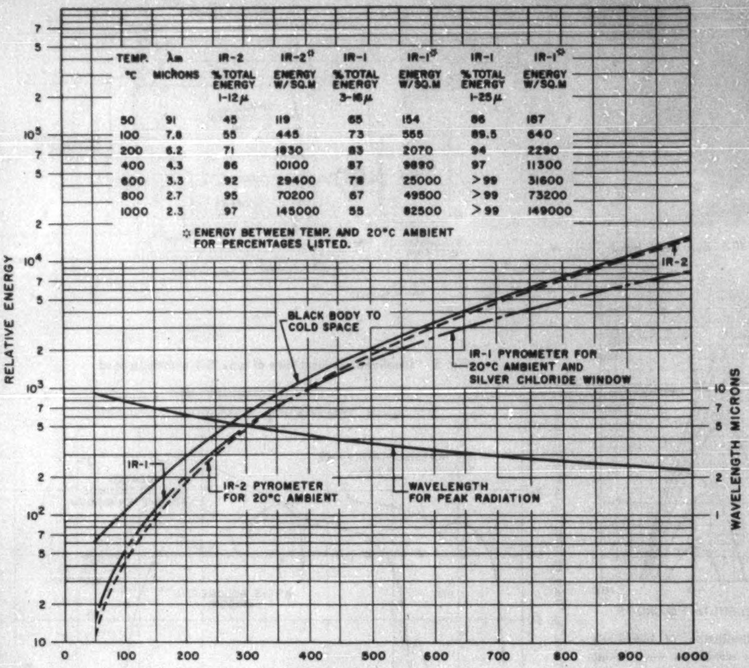
$$\text{(Wien)}\quad \lambda_m T = 2{,}900 \tag{3}$$

$$\text{(Stefan-Boltzmann)}\quad W = KT^4 \qquad W_1 = K(T^4 - T_0^4) \tag{4}$$

where

$W\lambda$ = radiant energy emitted per unit-area per unit-increment of wave length
$\left.\begin{matrix}W\\W_1\end{matrix}\right\}$ = radiant energy emitted per unit area
C_1 = first radiation constant
C_2 = second radiation constant
T = absolute temperature of radiating body, degrees Kelvin
T_0 = absolute temperature of surroundings, degrees Kelvin
λ = wave length in microns
λ_m = wave length in microns for maximum radiant energy
exp = 2.718, Napieran base
K = Stefan-Boltzmann constant

A family of curves of equation 2 was plotted in Fig. 6. Equation 3 gives the relationship of wave length to the temperature at which W_λ is maximum. It is convenient for rapidly estimating the



TEMP. °C	λm MICRONS	IR-2 % TOTAL ENERGY 1-12μ	IR-2☼ ENERGY W/SQ.M	IR-1 % TOTAL ENERGY 3-16μ	IR-1☼ ENERGY W/SQ.M	IR-1 % TOTAL ENERGY 1-25μ	IR-1☼ ENERGY W/SQ.M
50	91	45	119	65	154	86	167
100	7.8	55	445	73	555	89.5	640
200	6.2	71	1830	83	2070	94	2290
400	4.3	86	10100	87	9890	97	11300
600	3.3	92	29400	78	25000	>99	31600
800	2.7	95	70200	67	49500	>99	73200
1000	2.3	97	145000	55	82500	>99	149000

☼ ENERGY BETWEEN TEMP. AND 20°C AMBIENT FOR PERCENTAGES LISTED.

Fig. 8. Total energy received by pyrometer systems as compared to blackbody energy for various temperatures

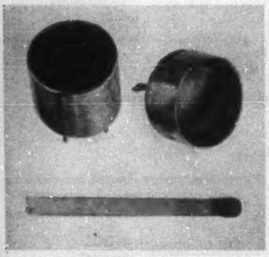


Fig. 9. Reference radiation source (left) and thermistor bolometer detector, as used in type IR-2 pyrometer head

energy distribution and is determined by setting the derivative of equation 2 equal to zero, and solving. Integration of equation 2 from zero to infinity with respect to λ gives the total radiated energy; see equation 4.

A further important property of blackbodies is the Lambert cosine law: The radiation from a blackbody for all wave length intervals varies as the cosine of the angle of emission.

As a consequence of this law and equation 4, the normal radiant intensity produced by a source of area A at a distance d is

$$P=5.67\times10^{-8}\,\frac{A}{\pi d^2}\,(T^4-T_0^4)\ \text{watts per square meter} \quad (5)$$

For a fixed field of view, the ratio of A to d is fixed, and for a 1-degree pyrometer field, equation 5 becomes

$$P=5.54\times10^{-12}\,(T^4-T_0^4)\ \text{watts per square meter} \quad (6)$$

Emissivity

The application of total radiation pyrometers to temperature measurement is based on equation 4 Since most radiating surfaces do not obey this law exactly, an emissivity correction factor is assigned

$$W_2=Ke(T^4-T_0^4) \quad (7)$$

This assumes that the radiation from T_0 is black, a condition which is approached in these pyrometers. Equation 7 gives the total energy radiated by a "gray" body; i.e., one which exhibits all of the blackbody properties except that its energy curve is displaced downward.

The emissivity is a function of temperature and, to define properly the spectral radiation, its variation with wave length must also be considered. Further, since ordinary bodies do not necessarily obey the cosine law, there is a variation with viewing angle. In radiation pyrometry, it is customary to use a single factor e_t, the total emissivity. This is an acceptable procedure if the assumptions involved are borne in mind.

Emissivities may be determined by a pyrometer such as the type *IR-1* with wide-band (1- to 25-micron) response. The instrument is calibrated against a known blackbody. Then, when aimed at the test material under a specified set of viewing conditions, the pyrometer will not indicate the true temperature T but some lower value T_1, termed the

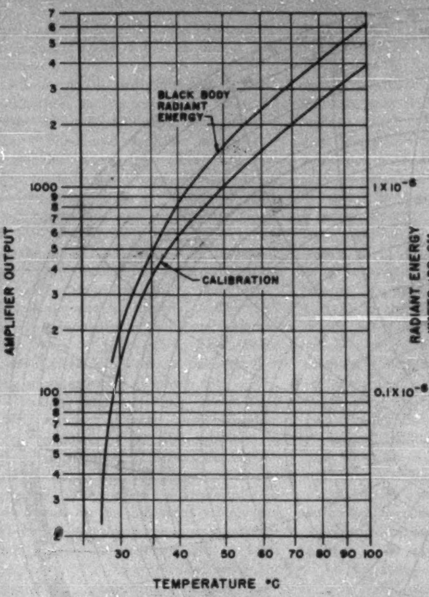


Fig. 10. Calibration curve for type IR-1 pyrometer at low temperatures

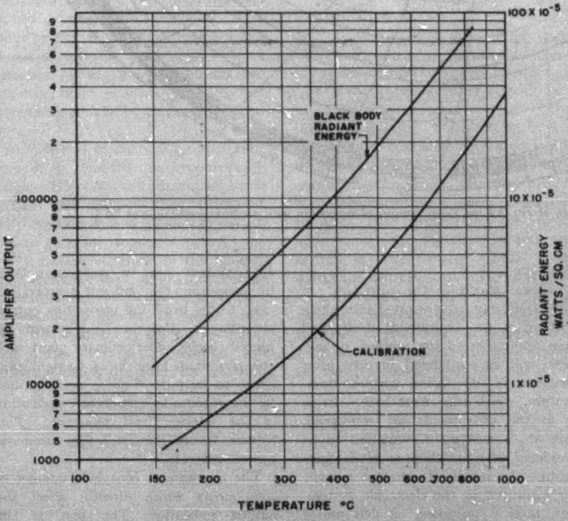


Fig. 11. Calibration curve for type IR-1 pyrometer at high temperatures

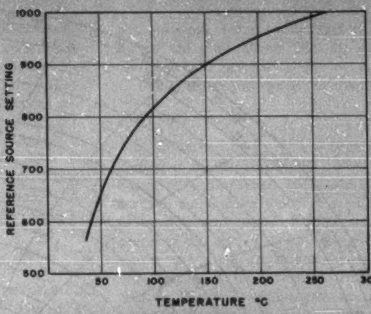


Fig. 12. Typical calibration curve for type IR-2 pyrometer with lens wide open. Higher temperatures are measured by reducing lens aperture

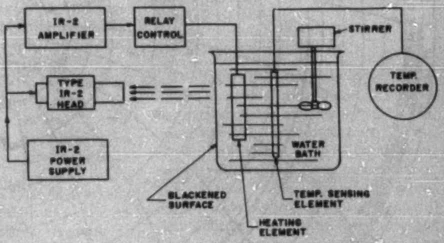

IR-2 AMPLIFIER
RELAY CONTROL
STIRRER
TYPE IR-2 HEAD
TEMP. RECORDER
IR-2 POWER SUPPLY
WATER BATH
BLACKENED SURFACE
TEMP. SENSING ELEMENT
HEATING ELEMENT

Fig. 13. Simple control system using type IR-2 pyrometer. Temperature-sensing element and recorder were used as measure of control efficiency

radiation temperature. For a non-blackbody

$$W_2 = K(T_1^4 - T_0^4) \tag{8}$$

Combining equations 7 and 8 gives, where T_0 is small,

$$e_t = \left(\frac{T_1}{T}\right)^4 \tag{9}$$

The total emissivity varies widely for different materials. At 700 C the value of e_t is 5% for polished silver, 10% for polished platinum, and 85% for iron oxide. At room temperature, the emissivity of bright copper is several hundred per cent less than at elevated temperatures. For clean metals, e_t is approximately representable by an equation of the form

$$e_t = BT^b \tag{10}$$

where B and b are constants.

For porous, nonmetallic materials,

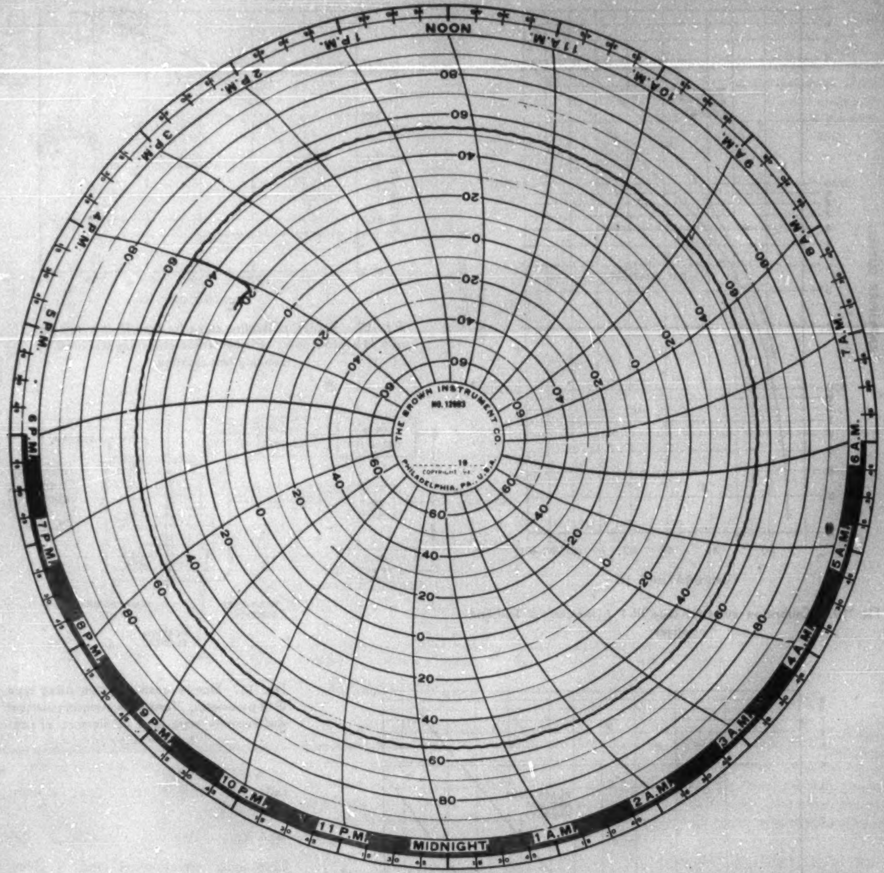


Fig. 14. Temperature chart for system of Fig. 13 for 1-day period of operation with control temperature at 54 C

e_t is high, regardless of the visible color of the material. Some nonmetallic materials have emissivities which decrease with increasing temperature. An additional consideration is the distinction between emissivity, which is the property of a material determined under standardized conditions, and emittance, which is the corresponding property of the surface. Thus, the published value of an emissivity may not be applied indiscriminately without consideration of surface condition and viewing angle, as well as temperature. For typical values of e_t see reference 1.

ACCURACY

Factors affecting the accuracy of calibration may be separated into those inherent in the instrument itself, and those applicable to the method of use. These will be considered in turn after discussion of one factor, ambient temperature, which can affect both. In so far as the instruments are concerned, there are a number of variables to be considered such as chopper blade temperature, variation of detector output, and variations in the electronic system. For practical purposes, as determined by measurements, the variation in the *IR-1* system may be compensated for by measuring the ambient temperature close to the head and computing output accordingly. For the *IR-2* system, repeated checks under ordinary plant variable conditions have shown no significant effect on output. This is accomplished by the automatic temperature regulation of the internal heat reference. This feature favors the *IR-2* for automatic process control.

The instruments contain a number of components which directly affect the output indication. The first is the collecting lens or mirror. These necessarily must be kept clean if the incoming radiation is to be held constant at the detectors. The latter are extremely stable chemically, with lattice structures, established by sintering at 1,300 C. After initial aging during manufacture, errors introduced by this source are negligible. Every reasonable effort has been made to stabilize the electric circuit and components so that frequent reset and calibration are unnecessary. An additional component in the *IR-2* system is the reference source. As indicated previously, its stability is dependent on a bead thermistor or platinum wire, inherently stable materials. The net result as indicated by preliminary laboratory tests is that calibration-checking once a week is all that is necessary for the most exacting measurements.

Factors external to the instrument itself are considerably more variable. The condition of the atmosphere is of major importance. Dirt, smoke, vapors, and humidity can all affect the output reading. The effect of ambient air conditions on the measured sample can be large. An object can have its surface temperature changed by five or more degrees by a slight breeze, with the pyrometer measuring surface temperature. Thus the ambient wind condition on the sample must be considered in measuring surface temperature. Allowances may be made for all of these effects provided the average proportions are kept constant and the distance from source to pyrometer is not varied. This allowance may be made either by a correction factor, or by calibrating and measuring under identical conditions, the latter being preferable wherever it is possible.

Humidity is of particular importance, since it is generally an uncontrollable factor. However, for distances under ten feet and ordinary plant conditions, the correction is very small and usually masked by other sources of error.[2]

The surface condition of the material to be measured is of utmost importance. Calibration should always be made on a representative sample, to avoid errors from this source. If the surface condition varies, several calibrations with different samples should be made to determine the extent of the possible error, and the corrections to be made. As a corollary to this factor, the angle of view should be considered, since the calibration will vary with this factor for most bodies.

A final precaution to be taken is the avoidance of stray radiation. This is a very important, often disregarded, source of error. Since an ordinary body is not a perfect absorber, it will reflect radiation from near-by objects. If a radiation source is in proximity to the source being measured, energy from it is likely to enter the pyrometer by reflection and to affect the indication. The filter-coated silver chloride detector window which cuts off radiation below three microns is helpful in eliminating stray radiation from visible sources such as lights, etc. An aperture plate limits the field of view on the type *IR-2* system. Additional radiation shielding should be used where there is any question of stray radiation affecting the readings.

One further consideration is of interest. In many continuous-flow processes, an absolute temperature reading is not needed. The essential requirement is to be able to reproduce a process reading over a period of time. For such conditions of operation, the type *IR-2* system will give entirely satisfactory service over long periods, without recourse to elaborate calibration schemes.

TYPICAL CALIBRATION

Type IR-1

Typical calibration curves for the type *IR-1* system are shown in Figs. 10 and 11. The low-temperature curve was taken by looking at the blackened side of a metallic container with a controlled temperature water bath. The high-temperature curve was taken by looking into the muffle of a small laboratory furnace. For comparison, the black-body energy for the temperatures concerned is also plotted. This is the difference between the total energy for a temperature and the energy of the chopper blade at ambient temperature for the 1-degree field. At ambient temperature, the amplifier output (except for residual noise) and the energy are zero. Except for a scale factor, the curves are quite close. Fig. 10 shows the threshold sensitivity to be about 1 by 10^{-7} watts per square centimeter.

Type IR-2

A calibration curve for the type *IR-2* system is shown in Fig. 12. This was taken on a blackened hot plate. The surface temperature was measured with a small thermocouple located at the center of the field of view. The calibration is for the full aperture of the collecting lens, and the standard reference source. Higher temperatures are read by stopping down the lens. With the reference set at 100 C, the average differential sensitivity is about ±5 C for full output. The incremental sensitivity is less than 0.25 C, and the over-all control accuracy of a typical installation is better than 0.5 C

APPLICATION

The type *IR-1* system is applicable to a wide variety of measurements such as the determination of chimney hot spots, line joint and transformer overheating, flame temperatures, emissivity measurements with regard to radiation losses, vacuum-tube plate dissipation where nonsinusoidal excitation makes computation difficult, and other determinations limited only by one's imagination. The efficacy of the type *IR-2* system to control applications is best illustrated by an example. Fig. 13 shows a simple control system in which the temperature of a water bath was sensed by the *IR-2* unit. The output of the unit was used to control input to the heater on an on/off basis, and a temperature-sensing element of the mercury type with recorder were used to measure the effectiveness of control. Fig. 14 is a typical temperature chart for the system of Fig. 13 and indicates the stability which is obtainable, even at the low control temperature of 54 C.

References

1. SMITHSONIAN PHYSICAL TABLES. Smithsonian Institution, Washington, D. C., 9th revised edition, 1954, Tables 81, 82, p. 100.

2. *Ibid.*, Table 567, p. 546.

Transistor Power Amplifiers with Switched Mode of Operation

A. G. MILNES
MEMBER AIEE

A CONVENTIONAL circuit for class *B* power amplification with transistors is shown in Fig. 1(A). The two transistors (p-n-p) are connected in the common emitter arrangement and in the absence of signal are quiescent at a point on the output characteristics such as *P* in Fig. 1(B). When a sinusoidal voltage signal is applied as base drive, the transistors operate on a load line such as *PQ* in alternate half-cycles, and a sinusoidal output is obtained. At any intermediate point on the load line *PQ* the active transistor sustains a collector-emitter voltage equal to the difference between the battery voltage and the instantaneous output voltage. The product of this collector-emitter voltage and the current flowing gives the instantaneous power loss in the transistor (excluding the small dissipation associated with the base drive). Simple calculations based on these considerations show that the maximum possible efficiency of this arrangement is 78 per cent and that maximum power dissipation occurs in the transistor when operating with an output that is 64 per cent of the possible full output voltage.

Power dissipation in the transistor with resultant temperature rise is the factor that principally limits the output power that can be handled by a given unit. In recent papers, however, the advantages of operating the transistor as a switch that is either of very high or of very low impedance, depending on the controlling conditions, have been recognized.[1,2] In the high-impedance condition, corresponding to, say, point *P* in Fig. 1(B), the transistor sustains voltage but there is little current flow and therefore low dissipation in the junction, and in the low-impedance condition the current flow is large but the collector-emitter voltage is small, as for point *Q* in Fig. 1(B), and therefore the dissipation remains low. The transition between *P* and *Q* is made rapidly and so represents very little heat generation within the transistor. Thus with switching circuit arrangements there can be appreciable uprating of the power-handling capabilities of a given transistor; for example, the values quoted for a typical power transistor are 10 watts in class *B* and 40 watts in switching applications.

Switching transistor power supply converters depending on these circuit principles have been described recently,[3,4] but possible applications of the technique to power amplifiers seem not to be discussed in the literature. A number of amplifier circuit arrangements based on this principle are therefore presented and discussed.

Half-Wave Switching Circuit with D-C Signal

As a simple form of the type of circuit to be examined, consider a p-n-p transistor connected between a load resistance and an a-c supply, as in Fig. 2(A), with a rectifier diode in series that ensures half-wave action. The base drive of the transistor is provided by the direct-voltage signal in series with an alternating bias voltage of the same frequency as the supply and the resistor R_b serves merely to restrict the base current to a safe limit under all input conditions.

A p-n-p transistor assumes a high impedance (cut-off) condition when the potential of its base is positive with respect to both the emitter and collector potentials. Thus, for the bias and signal voltages in Fig. 2(B), the transistor is cut off between the angular position θ and π-θ and the full supply voltage appears across the emitter and collector terminals as shown in Fig. 2(C). However, in the intervals 0 to θ, and π-θ to π, the base potential tends to be negative with respect to the emitter, and the transistor becomes conducting. Almost the full supply voltage appears across the load resistance in these intervals, providing the base drive current is adequate to permit the required load current flow (for example, from Fig. 1(B) a base current flow of 30 milliampere peak is required for a peak load current of 450 milliampere). In the second half-cycle, π to 2π, the supply voltage is blocked by the diode rectifier in the circuit. For the idealized circuit behavior thus described the cut-off angle is given by

$$\sin\theta=\frac{V_{d-c}}{V_{\text{bias}}} \tag{1}$$

and the mean load voltage by

$$V_{\text{load}}=\frac{V_p}{\pi}(1-\cos\theta) \tag{2}$$

where V_{bias} is the peak value of the sinusoidal bias voltage and V_p is the peak value of the supply voltage.

From equations 1 and 2, the theoretical output characteristics, Fig. 2(K), curve 1, can be calculated. If the bias voltage is of reversed phase with respect to the supply voltage (the dots shown in Fig. 2(A) indicate the in-phase condition), and if the signal voltage is negative, the load voltage wave form becomes of the type shown in Fig. 2(C) since excess of bias over signal voltage then produces the base polarity corresponding to conduction and not cut-off. The output characteristic that results is shown as curve 2 in Fig. 2(K). From curves 1 and 2 it is clear that an intermediate condition such as 90-degree lead bias is of interest. The wave forms for this are given in Fig. 2(F) through (J). The circuit equations are as follows:

$$\cos\theta=\frac{V_{d-c}}{V_{\text{bias}}} \tag{3}$$

$$\text{and } V_{\text{load}}=\frac{V_p}{2\pi}(1+\cos\theta) \tag{4}$$

thus, the resulting characteristic is the straight line 3 in Fig. 2(K).

If the bias voltage is between one-fifth and one-tenth of the supply voltage, the required signal voltage is similarly small; and since the base current is an order of magnitude less than the load current the characteristics of the circuit correspond to some degree of power amplification. The magnitude required for the bias voltage is determined by the base current required to permit flow of the peak load current and by the sharpness of switching required by dissipation considerations.

A d-c supply source is a possible alternative to the half-wave a-c source shown in Fig. 2(A): the output characteristic for this condition (somewhat nonlinear) may be readily calculated if required.

Paper 56-94, recommended by the AIEE Solid State Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted October 19, 1955; made available for printing February 24, 1956.

A. G. MILNES is with Carnegie Institute of Technology, Pittsburgh, Pa.

The author wishes to acknowledge his indebtedness to Prof. L. A. Finzi and colleagues at Carnegie Institute of Technology for the facilities provided under the United States Research Fellowship.

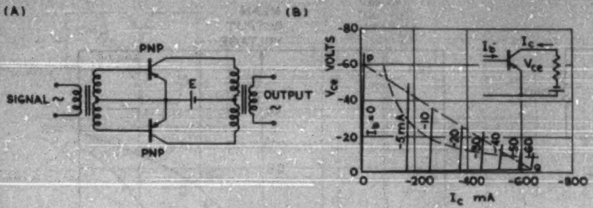

(A)
PNP
SIGNAL ~
E
OUTPUT ~
PNP
(B)
-80
V_{ce} VOLTS
-60
-40
-20
0
P
$I_b = 0$
-5 mA
-10
-20
-30
-40
-50
-60
Q
I_b
I_c
V_{ce}
0
-200
-400
-600
-800
I_c mA

Fig. 1. Conventional power amplifier with p-n-p transistors

A—Class B common-emitter connection
B—Typical output characteristics for power transistor

Half-Wave Switching Circuit with A-C Signal

Fig. 3 shows the behavior of the basic half-wave circuit controlled by an a-c signal of supply frequency.

For a bias voltage leading the supply by 90 degrees the idealized wave forms are as shown in Fig. 3(B) through (D). If the bias leads by ψ degrees the circuit equations are

$$V_{\text{bias}} \sin(\theta+\psi) = V_{\text{sig}} \sin\theta \qquad (5)$$

and $$V_{\text{load}} = \frac{V_p}{2\pi}(1+\cos\theta) \qquad (6)$$

Theoretical output characteristics obtained from these equations are given in Fig. 3(E) and it will be seen that for this a-c signal condition with a 90-degree lead bias there is loss of sensitivity at high and low output levels. Curve 3 for the in-phase bias condition does not represent true switching operation of the transistor since cut-off and conduction periods of the kind discussed previously do not occur. Therefore, if on curve 3 an output intermediate between full and zero is required, the transistor while conducting also sustains voltage, and so appreciable heat loss occurs which must be dissipated by the unit.

For the characteristics of Fig. 3(E) the signal voltage is either in phase with or in reverse phase to the supply. Characteristics for any other phase-angle relationship between signal and supply are easily derived.

Full-Wave Parallel Circuit with A-C Output

A full-wave output is obtained if two transistors active on alternate half-cycles are connected in parallel as shown in Fig. 4(A). For the signal voltage in phase with the supply and the bias leading by 90 degrees, the idealized wave form of load voltage is that in Fig. 4(C). (The arrows indicate that both triggering angles advance with increase of signal.)

The theoretical output of the parallel circuit is, of course, twice that for the half-wave arrangement and the characteristic (for a 90-degree lead bias) is similar to curve 1; see Fig. 3(E). An experimental characteristic for this condition is given as curve 1 in Fig. 4(D). The signal voltage plotted is the input to the signal voltage transformer which had two secondary windings, each of half the number of primary turns. The bias windings formed part of a similar transformer and the input was 10 volts, rms. Hence the switching and signal voltages actually operating in the transistor base circuits were of about 5 volts in magnitude. The resistors R_b limiting the base current flow were of 50 ohms. The supply voltage (60 cycles per second) was 50 volts rms and was perhaps somewhat too high for transistors with a 60-volt peak collector rating, although it gave satisfactory wave forms at room temperature.

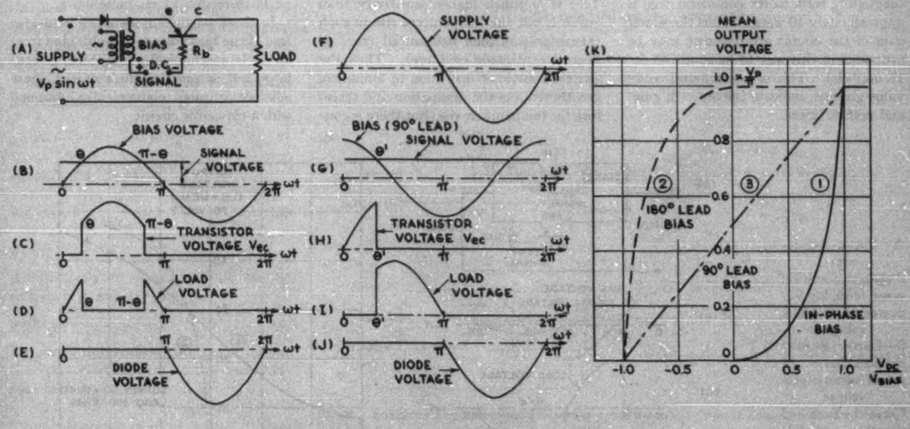

(A)
SUPPLY ~ $V_p \sin \omega t$
BIAS
R_b
D.C. SIGNAL
LOAD
e
c
(B)
BIAS VOLTAGE
θ
π-θ
SIGNAL VOLTAGE
ωt
0
π
2π
(C)
θ
π-θ
TRANSISTOR VOLTAGE V_{ec}
(D)
LOAD VOLTAGE
θ
π-θ
(E)
DIODE VOLTAGE
(F)
SUPPLY VOLTAGE
(G)
BIAS (90° LEAD)
SIGNAL VOLTAGE
θ'
(H)
TRANSISTOR VOLTAGE V_{ec}
(I)
LOAD VOLTAGE
(J)
DIODE VOLTAGE
(K)
MEAN OUTPUT VOLTAGE
1.0 × $\frac{V_p}{\pi}$
0.8
0.6
0.4
0.2
0
②
180° LEAD BIAS
③
90° LEAD BIAS
①
IN-PHASE BIAS
-1.0
-0.5
0
0.5
1.0
$\frac{V_{DC}}{V_{BIAS}}$

Fig. 2. Half-wave switching amplifier with d-c signal input

A—Circuit arrangement
B through E—Ideal wave forms for in-phase bias voltage
F through J—ideal wave forms for 90-degree lead of bias voltage
K—Theoretical output characteristics

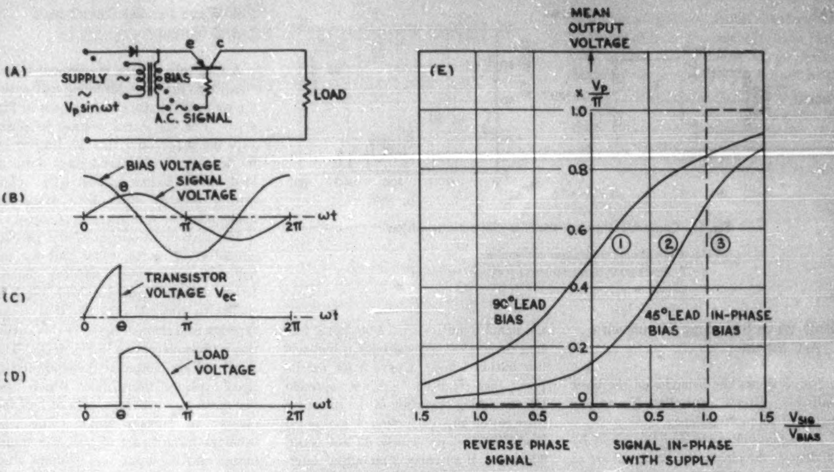

(A)
SUPPLY ~
$V_p \sin \omega t$
BIAS
A.C. SIGNAL
e
c
LOAD
(B)
BIAS VOLTAGE
θ
SIGNAL VOLTAGE
ωt
0
π
2π
(C)
TRANSISTOR VOLTAGE V_{ec}
0
θ
π
2π
ωt
(D)
LOAD VOLTAGE
0
θ
π
2π
ωt
(E)
MEAN OUTPUT VOLTAGE
$\times \frac{V_p}{\pi}$
1.0
0.8
0.6
0.4
0.2
0
①
②
③
90° LEAD BIAS
45° LEAD BIAS
IN-PHASE BIAS
1.5
1.0
0.5
0
0.5
1.0
1.5
$\frac{V_{SIG}}{V_{BIAS}}$
REVERSE PHASE SIGNAL
SIGNAL IN-PHASE WITH SUPPLY

Fig. 3. Half-wave switching amplifier with a-c signal input

A—Circuit arrangement
B through D—Ideal wave forms for 90-degree lead of bias voltage
E—Theoretical output characteristics

For the signal range of zero to 10 volts the input current to the signal transformer varied between 15 and 30 milliamperes; thus the input power was of the order of 1/4 watt. The change in power output of curve 1 between these signal limits (neglecting form factor considerations) is approximately 10 watts. Thus the power gain of the circuit arrangement may be estimated in this way to be about 40 (or 16 decibels). This is a rough-and-ready value given to indicate the order of gain, and nothing more.

With the circuit operating in the switched mode resulting in curve 1 of Fig. 4(D), the temperature rise of each transistor was quite small (a few degrees centigrade). With no bias applied for switching purposes, curve 2 was obtained. This is of much higher sensitivity than curve 1, but the temperature rise in each transistor was then substantial (say, 40 degrees centigrade or more). Thus the switched mode of operation in amplifiers greatly reduces the dissipation and therefore the temperature rise, but there is considerable loss in sensitivity. Switched operation has a second advantage in that the output curve is not so directly related to the immediate transistor parameters (which are temperature-dependent) and therefore is virtually a stabilized (or negative-feedback) characteristics.

In these particular tests the 90-degree lead of the bias voltage was obtained from an electromechanical phase-shifting regulator, but in practice the required phase advance is more conveniently obtained with a capacitor circuit.

Vol. 7
Pt. 1
1956

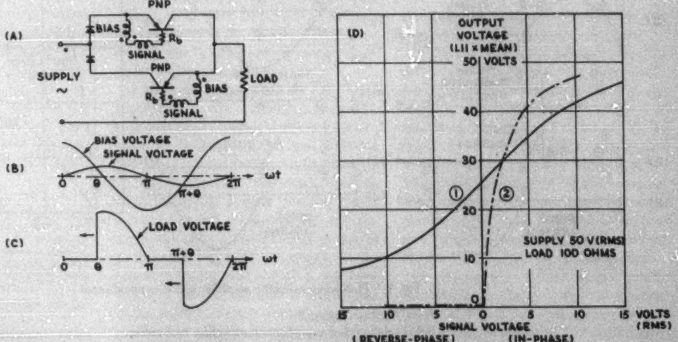

(A)
PNP
BIAS
R_b
SIGNAL
PNP
SUPPLY ~
BIAS
R_b
SIGNAL
LOAD
(B)
BIAS VOLTAGE
SIGNAL VOLTAGE
0
θ
π
π+θ
2π
ωt
(C)
LOAD VOLTAGE
π+θ
0
θ
π
2π
ωt
(D)
OUTPUT VOLTAGE (1.11 × MEAN)
50 VOLTS
40
30
20
10
0
①
②
SUPPLY 50 V(RMS)
LOAD 100 OHMS
15
10
5
0
5
10
15 VOLTS (RMS)
SIGNAL VOLTAGE
(REVERSE-PHASE)
(IN-PHASE)

Fig. 4. Full-wave parallel circuit with a-c signal and a-c output

A—Circuit arrangement
B through C—Ideal wave forms for 90-degree lead of bias voltage
D—Experimental characteristics of output versus signal voltage
Curve 1—Switched operation with 90-degree lead bias of 10 volts, rms
Curve 2—Linear operation with no bias

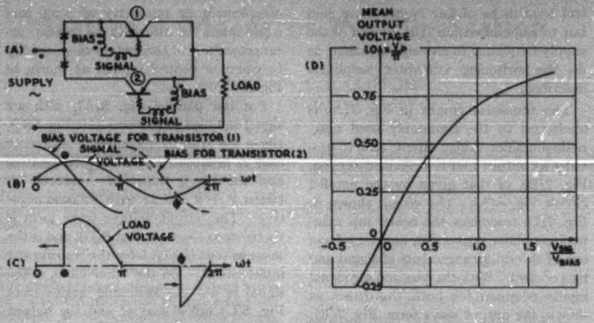

(A)
①
BIAS
SIGNAL
SUPPLY ~
②
BIAS
SIGNAL
LOAD
BIAS VOLTAGE FOR TRANSISTOR (1)
SIGNAL VOLTAGE
BIAS FOR TRANSISTOR (2)
(B)
0
θ
π
φ
2π
ωt
LOAD VOLTAGE
(C)
0
θ
π
φ
2π
ωt
(D)
MEAN OUTPUT VOLTAGE
1.0 = Vp/π
0.75
0.50
0.25
0
-0.25
-0.5
0
0.5
1.0
1.5
V_SIG/V_BIAS

Fig. 5. Two-transistor circuit with polarity of mean output changing with reversal of signal phase

A—Circuit arrangement, similar to that of Fig. 4(A) but with signal for transistor 2 reversed
B through C—Ideal wave forms for 90-degree lead of bias voltage
D—Theoretical characteristic for 90-degree lead of bias voltage

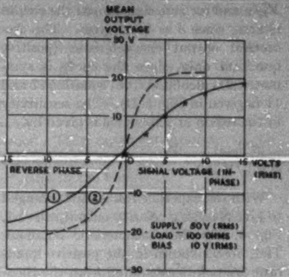


Fig. 6. Experimental characteristics for circuit of Fig. 5(A)

Curve 1—Switched operation with 90-degree lead bias of 10 volts, rms
Curve 2—Linear operation with no bias voltage
Points added to curve 1 calculated from theory

Two-Transistor Circuit with Polarity Reversal of Mean Output

The parallel transistor circuit of Fig. 4(A) has equal triggering periods in each half-cycle of operation, since a given signal sense causes either an advance or retard of switching in both half-cycles. Polarity-reversing characteristics are obtained, however, if the circuit is revised so that the switching angle advances in one half-cycle and retards in the other for a given signal sense. By the reversal of one signal winding as in Fig. 5(A), this effect is achieved. Fig. 5(B) illustrates bias and signal conditions that result in triggering angles θ and ϕ and these give an output voltage wave form as in Fig. 5(C).

For a 90-degree lead of the bias voltage the switching angles are given by

$$\cot\theta = \frac{V_{\rm sig}}{V_{\rm bias}} \qquad (7)$$

where $0 < \theta < 180°$

$$\text{and } \cot\phi = -\frac{V_{\rm sig}}{V_{\rm bias}} \qquad (8)$$

where $180° < \phi < 360°$

Alternatively to equation 8, it can be seen from symmetry that

$$\phi = 2\pi - \theta \qquad (9)$$

By simple integration, the mean output (i.e., direct component of the load voltage) is

$$V_{\rm load} = \frac{V_p}{2\pi}(\cos\theta + \cos\phi) \qquad (10)$$

$$= \frac{V_p}{\pi}\cos\theta \qquad (11)$$

The maximum output is therefore ±

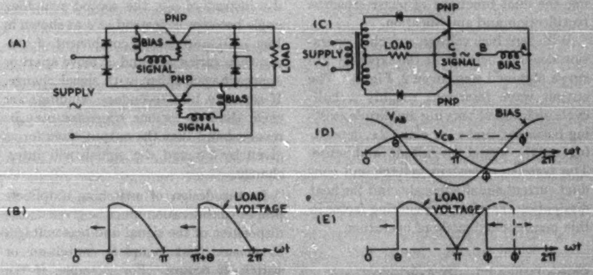

(A)
PNP
BIAS
SIGNAL
PNP
BIAS
SIGNAL
SUPPLY ~
LOAD
(B)
LOAD VOLTAGE
0
θ
π
π+θ
2π
ωt
(C)
PNP
SUPPLY ~
LOAD
C
~
SIGNAL
B
BIAS
A
PNP
(D)
V_AB
V_CB
BIAS
θ
φ
φ'
0
π
2π
ωt
(E)
LOAD VOLTAGE
0
θ
π
φ
φ'
2π
ωt

Fig. 7. Other switching amplifier arrangements

A—Bridge connection
B—Output of bridge connection
C—Center-tap circuit with common-emitter connection
D through E—Wave forms for center-tap circuit

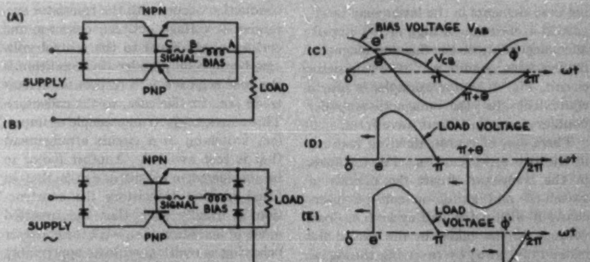

(A)
NPN
C
B
A
SIGNAL
BIAS
SUPPLY ~
PNP
LOAD
(B)
NPN
SIGNAL
BIAS
LOAD
SUPPLY ~
PNP
(C)
BIAS VOLTAGE V_AB
θ'
θ
V_CB
φ'
0
π
π+θ
2π
ωt
(D)
LOAD VOLTAGE
0
θ
π
π+θ
2π
ωt
(E)
LOAD VOLTAGE φ'
0
θ'
π
2π
ωt

Fig. 8. Switching amplifier circuits using n-p-n and p-n-p transistors in complementary symmetry

A—Circuit arrangement with single inputs for bias and signal voltages
B—Circuit with rectified output
C—Bias and signal voltage wave forms, 90-degree lead bias
D—Load voltage output of circuit A for a-c signal
E—Load voltage output of circuit A for d-c signal

V_p/π and for zero signal input the output is zero, since $\theta = 90$ degrees. The theoretical output characteristic (positive quadrant only, since the curve is symmetrical) calculated from equations 7 and 11 is given in Fig. 5(D). The sensitivity in the region of zero signal is given by

$$\frac{d(V_{\text{load}})}{d(V_{\text{sig}}/V_{\text{bias}})} = \frac{V_p}{\pi} \tag{12}$$

With the experimental circuit arranged to give reversing polarity output, curve 1, Fig. 6, was the characteristic obtained. The crosses shown in the positive quadrant are points calculated from the theoretical characteristic of Fig. 5(D) and these are in good agreement with the experimental curve. The gain of this arrangement was not high; for example, for a signal of 5 volts the input was about 0.1 volt-ampere and the output about 1.0 volt-ampere (mean). However it should be recognized that the circuit is performing the dual functions of phase-sensitive rectification and amplification.

With zero bias voltage, switched operation does not occur, and the experimental curve obtained was curve 2, Fig. 6. The output wave form was roughly a half-cycle sinusoid of varying amplitude existing between 0 and π, or π and 2π, depending on the signal magnitude and sense. The transistors sustain voltage and conduct current simultaneously, and the heat dissipation therefore is a limiting factor in this nonswitched mode of operation.

Some Other Circuit Arrangements

The basic switching amplifier concept may be applied in numerous circuit forms containing two or even four transistors. Many of these are analogous to magnetic-amplifier circuits since the saturating core elements in the latter may be regarded as switches of high or low impedance, depending on the flux and magnetomotive-force conditions. The transistor circuit, Fig. 4(A), for example, is nearly equivalent to the magnetic-amplifier doubler circuit in output wave form.

There are, of course, differing features in the two arrangements. For instance, in the transistor circuit the periods of switch-off and switch-on can be interchanged within the half-cycle (i.e., load voltage can be taken in the period 0-θ instead of θ to π) by reversing the signal and bias polarities. This possibility does not exist in the magnetic-amplifier doubler circuit for reasons in connection with the required flux changes.

The signal input to a magnetic amplifier is usually d-c but this is not always so.[5] Magnetic-amplifier stages with a-c control tend to be of low or moderate gain but to respond within half a cycle of the supply frequency to any signal change, as do the switching transistor amplifiers described.

The transistor circuit in Fig. 7(A) is analogous to the commonly used magnetic-amplifier bridge circuit and it provides an output that is a rectified version, Fig. 7(B), of that given by the parallel circuit Fig. 4(A). The circuit shown in Fig. 7(C) resembles the center-tap magnetic-amplifier connection and may have one of several arrangements of signal and bias control. With the bias and a-c signal inputs common for both transistors as shown, the output wave form, Fig. 7(E), is obtained. Since θ and ϕ advance or retard together with change of signal amplitude, the effect is equivalent to a phase shift of the a-c component of output with the d-c component remaining constant. On the other hand, if the signal is d-c instead of a-c, the second switching angle becomes ϕ' instead of ϕ as shown in Fig. 7(D), and the d-c component of output then varies, as θ and ϕ' move apart or approach each other, with signal change. If separate bias secondary windings are used, the bias for one transistor may be reversed and then the output wave forms given by a-c and d-c signals will interchange.

In the design of switching amplifiers proper consideration must be given to the disposition of the signal and bias voltages to ensure that complete switch-on or switch-off occurs. For example, if the collector in a p-n-p transistor is positive with respect to the emitter by the extent of the supply voltage, and if the control voltage is placed between base (positive) and emitter, switch-on occurs when the supply voltage exceeds the control. But conduction occurs with the transistor sustaining a voltage between collector and emitter about equal to the control voltage, because only under this condition is the base negative with respect to an electrode (i.e., in this case, to the collector). This corresponds to an example of imperfect switch-on in a circuit arrangement that is best avoided. Another factor to be considered in circuit design is that in fused-junction transistors the construction is usually such that the inverted mode of operation (i.e., with the collector behaving as emitter) requires appreciably more base current drive for a given output current flow.[2]

With two similar transistors, these principles tend to preclude the use of signal and bias inputs that are common for both units, and therefore input transformers are usually required. However, if the complementary symmetry of p-n-p and n-p-n units is utilized, the circuit arrangements lend themselves more readily to common control inputs, as shown in Fig. 8.

For the parallel Fig. 8(A), with a-c signal and bias wave forms as in Fig. 8(C), the n-p-n transistor conducts between θ and π, because its base is then positive and the p-n-p transistor conducts between $\pi + \theta$ and 2π with its base negative. The output voltage wave form is therefore as shown in Fig. 8(D), and is the same as that in Fig. 4 for the transformer input circuit. On the other hand, if the signal is d-c the switching angles from Fig. 8(C) are θ' and ϕ' and the output wave form is as shown in Fig. 8(E). This resembles the wave form of Fig. 5(C), produced in the simple p-n-p circuit by an a-c signal. The circuit in Fig. 8(B) is a bridge version of Fig. 8(A) and provides a rectified load output.

Conclusions

A switched mode of operation for transistor power amplifiers results in very low power loss in the transistor and therefore the temperature rise is small and the transistor retains appreciable power-handling capabilities up to high ambient temperatures. The characteristics obtained are of lower gain than for linear operation but are correspondingly less sensitive to changes in the actual transistor parameters. The switched output wave forms tend to resemble those of magnetic amplifiers, and certain circuits are analogous.

Within the scope of this paper it is not possible to discuss matters such as the influence of load on circuit behavior, circuits requiring more than two transistors, and circuit techniques for transistor stages preceding such amplifiers. However, the general concept appears promising and should repay continued circuit investigation.

References

1. Control Applications of the Transistor, **E. F. W. Alexanderson.** *Proceedings,* Institute of Radio Engineers, New York, N. Y., vol. 40, Nov. 1952, pp. 1508–11.

2. Junction Transistors Used as Switches, **R. L. Bright.** *AIEE Transactions,* vol. 74, pt. I, Mar. 1955, pp. 111–21.

3. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, **G. H. Royer.** *Ibid.,* July 1955, pp. 322–26.

4. Phase Locking of Switching-Transistor Converters for Polyphase Power Supplies, **A. G. Milnes.** *Ibid.,* Nov. 1955, pp. 587–92.

5. A-C Controlled Transductors, **A. G. Milnes, T. S. Law.** *Monograph no. 141M,* Institution of Electrical Engineers, London, England, 1955.

Heat as a Means of Air Purification

J. C. BECKETT
MEMBER AIEE

C. E. CLIFTON
NONMEMBER AIEE

Synopsis: These studies support the view that high-temperature heaters, such as the ceramic-type incandescent electric heater, exert a considerable bactericidal and sporicidal effect. When in operation in the home, the rates of reduction of air-borne micro-organisms generally are of the order of 50 per cent per hour for the more resistant spores. Where these electric heaters are the principal source of heat for a given home or room, it is reasonable to expect that the micro-organism content will reach a very low level after several hours of operation. It is clear that a considerable portion of the air circulating around the heater may not soon pass close enough for the organisms contained therein to be destroyed. It takes several hours of operation to achieve a stable minimum condition. The germicidal action appears to be much more rapid than that of ultraviolet light as ordinarily placed in a room but not so fast as with the use of a chemical material such as propylene glycol vapor.

MANY developments in recent years have given cause for concern about air contamination. The advent of irritating smog in urban areas, the threat of biological warfare, and the increasing awareness that air contamination is a health hazard have stimulated thought on this subject. Mechanical filters of various sorts have been used rather successfully but too few people are aware that the proper application of heat can be effective. The use of heat to purify the air may seem new to some, but it has been applied for many years in the exhaust-air systems of bacteriological laboratories.

Air purification in the home should not involve the control of a large volume or weight of objectionable material. Ideally, domestic air-cleaning equipment must be simple; it must have a low first cost, low operating cost, no significant maintenance cost, and require no regular inspection. It is logical to consider air-purification equipment as part of the heating or air-conditioning system rather than as a separate piece of equipment, though it may be separate when more convenient.

Objectionable Contaminants

Of the many contaminants present in the air in homes, by far the most common and potentially the most harmful are living organisms such as mold spores, bacteria, and viruses. Toxic chemicals seldom occur and, because of the small quantities encountered, are primarily dangerous only because they may weaken resistance to infection.

Among these toxic chemicals, the most common in the home is ozone. There has been much confusion regarding the hazard of ozone. Though it may have limited usefulness in the control of objectionable odors, it is generally to be avoided,[1] and is recognized by medical authorities to be a health hazard when more than 0.10 ppm (part per million) is present.[2] In Los Angeles 0.80 ppm (under severe smog conditions) has been authoritatively reported by the Stanford Research Institute.[3] Smog alerts are proclaimed whenever the ozone concentration exceeds 0.50 ppm.

Other common harmful contaminants include pollens and ordinary house dust. It is well known medically that some persons are sensitive to pollens and to house dust (particles of wool-down, cotton, etc.). Pollens are relatively large in size, many being visible without the aid of magnification; they are easily filtered from the air. However, both pollens and house dust, or lint, can be destroyed by heat if the heat is of sufficient intensity and the time of exposure relatively long. These particles generally have adequate mass to impinge on the hot surfaces of the heater and incineration has been visually observed.

Existing Methods of Air Purification

There are three available means of controlling air-borne contaminants in use today. These are well developed commercially and are reviewed here briefly.

OIL AND PROPYLENE GLYCOL

It is well known that contamination of the air in living areas is markedly affected by the dispersion of organisms during sweeping, bed making, and similar activities. The spreading of certain oils on floors quickly results in a reduction of the atmospheric bacterial count, reductions of 80 per cent having been reported.[4] Though this is an effective means of control of air-borne contaminants, it is not a welcome one to the home owner. This expedient appears to be both inconvenient, and, to a degree, impractical in the home.

Although propylene glycol has been used effectively and its lethal effect on bacteria and viruses is well known, it, too, is very inconvenient for use in the home. No one desires to maintain a vaporizer continuously generating glycol vapor, because of the humidity problem and the condensation that takes place. This means of air purification, therefore, is not practical except in unusual circumstances.

ULTRAVIOLET RADIATION

An accurate evaluation of ultraviolet radiation to control air-borne organisms as compared with heaters of the type described here is difficult because of the many variables that exist. A heater designed for the purpose has the distinct advantage over ultraviolet lamps placed along the upper part of the walls in that the heater circulates a substantial volume of air, whereas the lamps do not. Such a heater, placed close to the floor, has a better chance to intercept contaminants or secondary aerosols stirred up by walking. The lamps, on the other hand, located high for safety, will have little effect on secondary aerosols. They have the further disadvantage that, if powerful enough to do an effective job in a room in the home, they will also generate ozone. Ozone cannot be tolerated where the background is already too high because of urban air pollution. There is also the problem of visible light that might interfere with sleeping and the hazard to the eyes of the wall-type lamps when they are not mounted high enough.[5]

Ultraviolet lamps undoubtedly have their proper application in air purification—as in hospitals and public places where a good engineering job of placement and ventilation can be obtained. They can be used at all times of the year, whereas an electric heater is limited to cold-weather use. Limitation of the electric heater to wintertime is not important so far as home use is concerned when it is considered that most respiratory infections occur during the cold months and that ventilation, an excellent means of purification except in smog areas, is generally adequate in summer for the control of bacteria and viruses in the air.[6]

Paper 56-236, recommended by the AIEE Electrical Techniques in Medicine and Biology Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted November 17, 1955; made available for printing February 28, 1956.

J. C. BECKETT is with the Wesix Electric Heater Company, San Francisco, Calif., and C. E. CLIFTON is with Stanford University, Stanford, Calif.

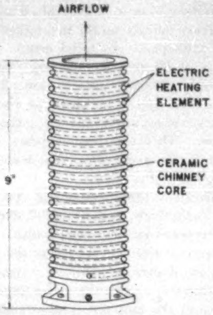


Fig. 1. Electric heating element showing hollow chimney-type ceramic core with resistance wire coil

Recently, attempts have been made to use ultraviolet radiation in ventilation systems in large buildings.[7] It was noted in this study that the percentage of destruction of bacteria (coli) was greater with the heater as compared with reported values with lamps, except when the microwatts per second per square centimeter of the lamps approached high levels.

ELECTROSTATIC AIR CLEANERS

The use of electrostatic devices for air cleaning is also well known. There are two main objections to this means of purifying air: 1. the first cost of the equipment is high, placing it out of reach of most home owners; and 2. there is always the problem of re-entrainment of dust and micro-organisms after collection. Where a regular program of cleaning of the places can be instituted, or where expensive automatic means for periodic plate cleaning is acceptable, this is an entirely satisfactory solution. However, it must be kept in mind that this is essentially a collection device and does not destroy the objectionable material. Proper use of heat positively assures that the air-borne contaminants are destroyed. This was an important consideration when an incandescent brick chimney was selected for the Naval Bacteriological Laboratories at Oakland, Calif., as the only known safe method for disposing of air-borne pathogenic organisms from ventilation exhaust systems.

Associated with electrostatic air cleaners is the possibility of introducing ozone into the air, which is an unacceptable risk in urban areas afflicted with air pollution. There is also the question of creating a positive space charge in a room supplied with air from an electrostatic air cleaner which is not ion-controlled, with the adverse effects this may have on asthma and hay-fever sufferers.[8]

Recently, an improved type of electrostatic air cleaner has been developed which incorporates a mechanical filter for better collection of the air-borne contaminants. This new development has many advantages but it has the same inherent disadvantage for domestic application that is inherent in all electrostatic and mechanical type filters; i.e., a regular maintenance or cleaning program is required to maintain a reasonable efficiency of the filter. In the home such a program is seldom practical and the cost of replacement filters is often excessive.

Test Results with the Use of Electric Heat

Various means of providing heat to sterilize air suggest themselves. One means is to provide a tunnellike oven through which air is passed at a controlled rate, allowing the air-borne particles to reach a temperature sufficient for their destruction.

A commercially available electric space-heating device consisting of a hollow ceramic core with an electric heating element surrounding the ceramic was selected for tests to be reported here. To be reasonably successful, it seemed essential to have incandescent heat on the inside of the ceramic core or chimney. Other means than electricity could have been used, but the latter was readily available and appeared to be the simplest where temperatures in excess of 1,000 degrees Fahrenheit (F) were desired and where the products of combustion of fuels must not contaminate or mix with the purified air. A device of this type is illustrated in Fig. 1. The ceramic core is 9 inches long by 3-inch outside diameter and $1^3/_4$-inch inside diameter. In the vertical position, approximately 10 cubic feet per minute circulate by natural convection. There is a marked advantage in having the air circulate through the ceramic chimney, wherein the heat surrounds the particles in the air, to neutralize the thermal forces that would otherwise drive the particles away from the source of heat. The open-coil electric element is covered by a charged grid to control the ions emitted from the hot element.[9]

In the tests described here, single-core Wesix ion-controlled heaters were used, some rated 1,000 watts and others 1,500 watts. The air flow averages 180 to 200 lineal feet per minute through the chimney core. The heating element, energized 30 minutes or more, operates at 1,200 to 1,500 F (measured with an optical pyrometer) and produces a temperature on the inside of the ceramic core of approximately 900 to 1,200 F.

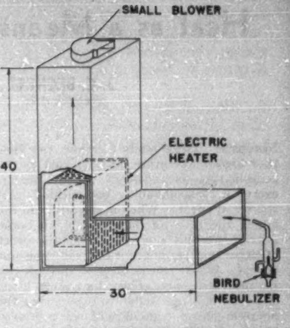


Fig. 2. Test arrangement for continuous nebulizing through electric heater

TESTS OF GERMICIDAL ACTIVITY

Three different test procedures were used:

1. Direct—in which a continuous stream of suspended organisms from a nebulizer was passed through and around the incandescent core of the heater.
2. Indirect—in an experimental room in which the concentration of the aerosolled suspended organisms could be controlled to some extent.
3. Indirect—in a residential room under natural conditions and the usual microbial content.

These three sets of tests were selected to test the effectiveness of a conventional electric room heater, as presently constructed, in the destruction of air-borne contaminants and to obtain general basic information that might lead to improvement in heater design for increased efficiency in air purification.

Direct—Using Continuous Nebulizing

In this set of experiments a Bird-type nebulizer was used (suspension in nebulizer being continuously replaced). The aerosol was directed into the open end of of an L-shaped wooden box and through a 1,000-watt electric heater in the corner of the L; see Fig. 2. Baffles were included so as to direct all of the air through and around the chimney core of the electric heating element. Provision was made to circulate air at a uniform rate through this device by means of a small fan, whether or not the heating element was energized. This test was designed to determine the degree of destruction when known particles passed through the chimney core at approximately operation speed and temperature.

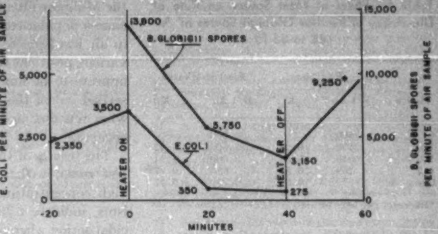
Samples of the effluent air were passed through a bubbler. This device consists of a standard glass flask arranged so that air to be sampled is drawn through a glass capillary tipped tube extending almost to the bottom of the flask and is impinged on a thin film of the solution; thereby assuring that the contaminants are retained in the solution and not carried on to the vacuum pump.[10] The volume sample is determined with the aid of a flowmeter. Experiments were conducted with viable spores (*Bacillus globigii*), vegetative bacteria (*Escherichia coli*), and an antistaphylococcus bacteriophage (Ph 65). Colony or phage counts were made in all tests by ordinary dilution and plating techniques and with the aid of a stereoscopic microscope.[11] In a few tests the relatively new Millipore filters[12] were tried with parallel results; however, under conditions of wide variation in numbers of organisms sampled, the Millipore filters were not as flexible as the usual dilution and plating technique.

In a preliminary test, plates of nutrient agar were exposed over the exhaust or chimney of the heater box. The results of this test were of interest in that a much higher colony count was obtained when the heater was in operation than when it was turned off. Later tests showed that this oddity was a result of rapid evaporation of water from the droplets coming from the nebulizer, many individual spores being set free. Since plates exposed directly to the exhaust did not give counts indicative of the total number of spores, but instead each colony developed from the spore or spores contained in separate droplets, this testing technique was entirely unsatisfactory.

This information proved useful, nevertheless; in experiments to be reported later, secondary heated air was introduced to the reflux chamber of a Wells-type nebulizer to reduce by means of evaporation the size of the aerosol droplets. This technique made it possible in experiments with bacteriophage to achieve particle size in the order of the size of the phage itself; i.e., 0.05 to 0.07 micron.

Tests using plate counts made from broth in impinger flasks (bubblers) indicated a marked reduction in numbers of viable spores as a result of the passage of spore-containing droplets through the heater. Fig. 3 illustrates typical tests and results with spores. Vegetative forms being much more susceptible to the lethal action of heat, the reduction is even greater than with the bacterial spores. *Bacillus globigii* spores are approximately 0.7 × 1.0 micron and roughly ellipsoidal in shape. They are much more heat resistant and more resistant to being nebulized and suspended in air than are vegetative bacteria. These spores proved to be the most convenient for experimental purposes. It seems reasonable to conclude that, if these spores can be destroyed by heat, mold spores of comparable size will also be destroyed, and bacteria and viruses even more rapidly.

Fig. 3. Viable counts of bacillus globigii spores and escherichia coli in aerosols (continuous spray) after passing through the heater, Fig. 2



*—Effluent air still quite warm but count approaches the original value

It should be noted that the moisture droplets containing the organisms coming from the nebulizer afford some degree of protection as the droplets go through the heater, with the result that the lethal action here is probably less effective than it might be with air-borne organisms under normal conditions. Also, the rate of airflow through the heater was slightly greater than that induced by the heater under normal use. Under the foregoing test condition, the percentage of spores in the airstream was reduced by 50 to 75 per cent, of bacteria by as much as 90 per cent, and of phage by about 99 per cent.

Indirect—Experimental Room with a High-Density Relatively Stable Microbic Population

The preceding tests gave a good indication of the potential germicidal activity that might be of practical use with a conventional type of electric heater. In the application of heat to a room it is not usual that all of the air supplied to the room be directed through the heater prior to its entry into the room. Rather, a high proportion of air is recirculated through the heat source so as to conserve heat, and it is problematical whether all of the air in the room will find its way through the heater or whether portions will be stagnant in corners of the room. Thus, purification by this means is to a great degree a dilution process whereby the content of microbes in the air is greatly reduced but never reaches zero. With this in mind, a second series of tests were conducted in a relatively small room, 350 cubic feet, in which the heater was permitted to circulate air in the normal way. In addition, a relatively high air-borne population of bacterial spores was aerosolled into the test room, with the use of a Wells-type nebulizer,[13] and the suspension was allowed to stabilize by standing.

In these tests, a 1,500-watt heater was selected with the thermostat either bypassed or set at its highest position, approximately 95 F. This was done because a 1,500-watt electric heater is a more popular size for home heating; and it produces higher temperatures than the smaller capacity heaters. Its thermostat was set high to keep the heater energized a reasonable length of time, considering the small size of the room. Control tests were performed under identical conditions with an electrically heated hot-water radiator to provide approximately the same room temperatures as obtained with the test heater itself.

The spores were sprayed into the room late in the afternoon so that a relatively stable population could develop over a 16-hour period. The rate of decrease in numbers of stabilized air-borne spores can be expressed by the equation

$$K = 2.303/t \log B_0/B_t$$

where B_0 is the number of viable organisms at the beginning of the time interval t and B_t is the number at the end. The velocity constant for the die-away was around 0.1 in the stabilized suspensions at 72 F; and about 0.8 to 0.9 in the higher temperature range when using the hot-water radiator. The calculated velocity constants were much higher when the radiant electric heater was in operation, being in the neighborhood of 1.25 to 1.50.

In similar tests, lower average room temperatures, 82 to 85 degrees, were maintained with the aid of a thermostat

Table I. Effect of Heat Source on Rate of Die-Away of Bacillus Globigii Spores in Air (82 to 85 F)

	Radiator		Electric Heater	
Hours	B/1 Air*	K	B/1 Air*	K
0	3800		20,000	
1.5	2500	0.28	10,600	0.42
3.0	1040	0.58	2,000	1.11
4.5	289	0.85	630	0.77
0 to 4.5		0.57		0.77

*Bacteria per liter of air.

control. *K* values near 1.0 were noted when using the electric heater as compared with 0.4 to 0.5 with the control radiator. In this test, the electric heater operated for only a few minutes at a time, the hollow core of the heater never reaching the maximum temperature.

Room temperature in excess of 80 F was found to be particularly detrimental to vegetative microbes but increased the rate of die-away to a lesser degree with *Bacilius globigii* spores. Use of the hot-water radiator for controls to obtain room temperatures equal to those observed with the electric heater removed any doubt about the results of this series of tests; see Table I.

It should be noted that considerable difficulty was experienced in maintaining stable suspension of relatively high microbic density. Air turbulence such as from a fan or air conditioner rapidly reduced numbers of suspended spores or vegetative cells. This is not to suggest that an air conditioner or fan has lethal effects but rather that, in an artificial environment as used here, these disturbing factors caused the spore-containing droplets or particles from the nebulizer to precipitate or impinge onto solid objects that might be encountered. These organisms are apt to become a source of secondary aerosols and lead to confusion in test procedure. Care had to be taken to maintain air movement in the test room at the level that might be encountered in the home. A commercial air conditioner (for the purpose of maintaining a room temperature of 72 degrees at all times with the heater in operation) could not be used in this experiment for the reason cited in the foregoing.

The percentages of spores settling on nutrient agar exposed in petri dishes in the test room were of a low order and of the same magnitude for either source of heat. There was less consistency in the results obtained with spores settling on petri dishes than with colony counts obtained from air samples either passed through broth in the bubbler or through the Millipore filters. This was interesting since a popular test for germicidal activity in air has been to expose petri dishes at various positions in a room. Such a test appears to be limited; it tells only what settled out of the air by gravity and not what remains air-borne to be inhaled by occupants.

The results of these tests in the special test room, with the use of a relatively high concentration of aerosolized organisms, indicate a highly significant sporicidal action when an electric heater of the open-coil radiant type is used as compared with a low temperature source, i.e., a hot-water radiator. The conditions being impractical for extrapolation to the home, it was decided to extend the tests by sampling air in a typical residence.

Indirect—In a Residence Room Under Natural Conditions

Tests in a home were conducted on widely different days and in different rooms but under normal living conditions. The numbers of bacteria per unit volume of air showed considerable variation, though the numbers were never high, but they were quite consistent in replicate samples collected the same day. This low density of organisms increases the error of measurement, requiring a large number of samples. A brief summary of these results has been reported.[14] A complete description follows:

Several samples of air were tested prior to the 1,500-watt Wesix heater being turned on. These agreed closely with each other for a given day, though large samples, up to 100 liters of air, were required. Results are illustrated in Table II. The percentage reductions (including higher fungi developing on incubation for 72 hours) observed after 1 hour of automatic heater operation (69 to 72 F) ranged from 46 to 56 per cent and in the following hour 30 to 80 per cent. Position of the heater in the room and the different air circulation patterns established probably account for the wide variations observed in longer tests.

Measurements in the home presented many difficulties primarily because of the low density of organisms as compared with the previous artificial conditions. As a precaution, three different methods of sampling were used. These were as follows:

1. Filtration of a measured volume of air through a Millipore filter, moistening the filter with nutrient broth, and transferring the filter to a nutrient pad for incubation.

Table II. The Effect of Wesix Electric Heaters on the Microbial Population of Air in the Home under Normal Conditions of Use. Temperature Range of 69 to 72 F[9]

		Colony Counts		
		Hours		
Location	Air Sample, Liters	0	1	2
Bedroom	40	88	44	8
Bedroom-study	35*	188	99	
Dining room	50	25	11	7
Dining room	100	74	43	31
Dining room	†	54	23	8
Dining room	100	16	8	3
Dining room‡	100	66	19	6
Dining room	†	68	21	14

*Millipore filters used for assay.

†Numbers settling on agar in petri dish in 15 minutes.

‡Wall-type heater instead of floor-model heater.

2. Filtration of a measured volume of air through broth in an impinger vessel and filtering the broth through a Millipore filter, the filter then being transferred to a nutrient pad.

3. Same as 2 except that aliquots were transferred to petri dishes and mixed with nutrient agar before incubation.

The Millipore filter was advantageous for filtering because of the small number of organisms present, but this was offset to a considerable degree because of the wide variations found in different rooms and at different times of the year (experiments in the home were conducted in December, January, and March), and the Millipore filter does not permit subsequent dilution to reduce numbers. In general, the bubbler technique of impingement in solution proved to be the most convenient method of obtaining an air sample.

Conclusions

Electric heaters of the incandescent type will reduce the microbial content of room air during the heating season by as much as 50 per cent per hour or better. The results of the direct aerosolation through and around the electrically heated ceramic core indicated that most of the destructive effect takes place inside the chimney core. Velocity of the particles moving through the core, temperature of the core, and the mass of the organic particles are principal factors. A longer period of exposure to the intense heat by either increasing the ceramic temperature or the length of the ceramic would improve performance. It was observed that very little destruction occurred in the air directed over the exterior surfaces of the ceramic core or the heating element. This is attributed to the fact that thermal forces acted to direct the particulate matter away from the hot body before destruction could be effected. It appears feasible that an electrostatic field could be used to overcome this thermal force and, thereby, enhance destruction through better utilization of hot surfaces on the exterior of the core.

For year-around air purification, a device using incandescent electric elements in combination with an electrostatic field to improve the efficiency, and with a cooling device to control the final air temperature, appears feasible, though perhaps not presently attractive economically.

References

1. Ozone Toxicity, **H. F. Stokinger.** *Archives of Industrial Hygiene & Occupational Medicine*, American Medical Association, Chicago, Ill., vol. 9, May 1954, pp. 366–83.

2. Heating, Ventilating and Air Conditioning Guide. American Society of Heating and Ventilating Engineers, New York, N. Y., vol. 33, 1955, p. 159.

3. The Smog Problem in Los Angeles County. Western Oil & Gas Association, Los Angeles, Calif., Jan. 1954, p. 11.

4. Air-Borne Infections, **A. P. Krueger**, et al. *War Medicine*, American Medical Association, Chicago, Ill., vol. 4, July 1943, p. 22.

5. *Ibid.*, pp. 20–23.

6. *Ibid.*, pp. 23–25.

7. Disinfecting Air with Sterilizing Lamps, **R. Nagy, G. Mouromsett, F. H. Rixton.** *Heating, Piping and Air Conditioning*, Chicago, Ill., Apr. 1954, pp. 82–87.

8. Artificial Air Ionization in Physical Medicine, **I. R. Kornblueh, J. E. Griffin.** *American Journal of Physical Medicine*, Baltimore, Md., vol. 34, no. 6, Dec. 1955, pp. 618–31.

9. Air Ionization as an Environment Factor, **J. C. Beckett.** *AIEE Transactions*, vol. 73, Sept. 1954, pp. 161–65.

10. Experimental Airborne Infection (book), **T. Roseburg.** Williams & Wilkins Company, Baltimore, Md., 1947, p. 222.

11. A Stereoscopic Method for Counting Bacterial Colonies, **C. E. Clifton.** *Journal of Bacteriology*, Baltimore, Md., vol. 69, no. 1, Jan. 1955, p. 107.

12. A Basic New Tool in Bacteriology, **J. J. Bush.** *The Scientific Monthly*, Washington, D. C., vol. 75, no. 5, Nov. 1952, pp. 303–05.

13. An Apparatus for the Quantitative Study of Airborne Respiratory Pathogens, **W. R. Lief, A. P. Krueger.** *Journal of Infectious Diseases*, Chicago, Ill., vol. 87, 1950, pp. 103–16.

14. Germicidal Activity of Electric Heaters, **C. E. Clifton.** *Science*, Washington, D. C., vol. 122, Oct. 21, 1955, p. 762.

The Pulse Stretch Coupling Circuit

H. W. PATTON
ASSOCIATE MEMBER AIEE

Synopsis: This paper describes a new coupling circuit applicable to both half-wave and full-wave magnetic amplifiers, particularly those amplifiers simultaneously requiring fast time response and large power gains. The pulse stretch technique uses a pulsating bias signal derived from an external supply to block the power winding of the reactor during reset. This blocking action permits free core control yielding an amplifier with an improved figure of merit.

General

MODERN high-performance equipments frequently require magnetic amplifiers having both fast response and large power gains. These requirements are typical of aircraft hydraulic control valves, instrument positioning servos, and other similar devices.

An early form of high-speed magnetic amplifier was reported by Ramey,[1] as shown in Fig. 1.

When operated in class B, E_b is adjusted to cause minimum current flow in the power circuit. Control current provided by E_c flows in mixing resistor R_m during the core reset cycle. This added current resistance drop through R_m reduces the net reset applied to the core, causing it to saturate earlier during the next firing cycle. E_c may be alternating current, direct current, or pulses, but to be most effective this control signal should match the wave form of E_b. To cause operation and cutoff with $E_c = 0$, E_b, R_m, or the reset winding turns may be adjusted. Usually, E_b is chosen with relation to the available E_c. If E_c is pulsed, as from a preceding magnetic amplifier, and E_b and E_p are sinusoidal, difficulties may be experience in achieving full control of the output core.

Another amplifier circuit reported by Ramey[1,2] is depicted in Fig. 2. This amplifier operates at full output with zero control. As E_c increases, assuming proper phase relationships, core reset is increased, gradually reducing the amplifier output.

For optimum operation and full control range, E_c should be of the same wave form as E_p and under maximum reset conditions the instantaneous value of E_c times the reactor turns ratio essentially equals E_p. If E_c exceeds this value, the voltage applied to the power winding rectifier will reverse, causing conduction in the power winding. This reflects a low impedance to the control winding compared to normal values of core-magnetizing current. For purposes of this paper the phenomenon just described will be termed "oversetting."

If E_c is pulsed, as from a preceding magnetic amplifier, difficulties may be experienced in obtaining proper amplifier control, as will be explained later.

For a maximum performance amplifier the type and amount of core material should be selected according to the available signal energy. For ordinary amplifiers this poses no particular problem. It is possible to construct an amplifier having an optimum quantity of core material so that the input energy under full signal conditions is just adequate to supply the B-H volume iron loss and the small copper loss. If the input energy is exceedingly small, the volume of iron called for may be less than 0.0001 cubic inch. Unfortunately, while such a design is constructable its manufacture is quite inconvenient and impractical. A typical reactor of this size may need 10,000 turns of no. *46* wire. If a larger core volume than the indicated amount is used, the input energy will be insufficient to control the core over more than a small portion of the firing angle. Here is the crux of the problem. Assuming the input energy is very small, so that it is impractical to use a core having an ideal configuration, and a larger core is used, the input energy is found only sufficient to control the core over a small part of its normal 180-degree range. This means that the output signal from the first core is a spike which occurs during the last part of the power cycle. If coupling the output from this core to another half cycle stage is attempted, it is discovered that the bias voltage times turns ratio requirement must still be satisfied. Matching on a proper voltage basis reveals that the output stage is controlled only over the same small firing angle as the input core. On the other

---

Paper 56-30, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted June 7, 1955; made available for printing March 29, 1956.

H. W. Patton is with the Airpax Products Company, Baltimore, Md.

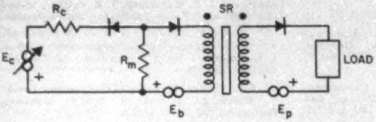


Fig. 1. Voltage-controlled amplifier with mixing resistor

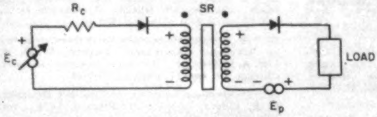


Fig. 2. Voltage-controlled amplifier

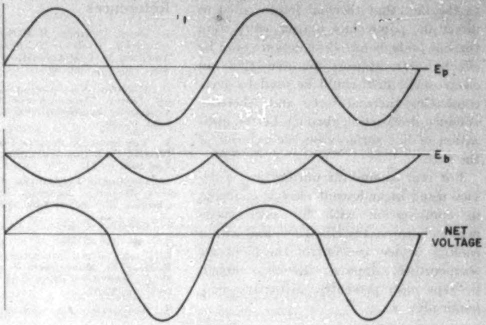


Fig. 4 (above right). Pulse stretch amplifier wave forms

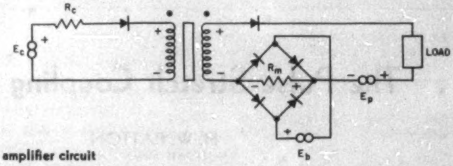


Fig. 5. Pulse stretch amplifier circuit

hand, if achieving a step-up ratio is tried, control energy is lost by circulation in the power winding through overset.

If a very large amount of control energy is available it may be possible to overpower the core despite this circulating current in the power winding. Unfortunately, these operating conditions do not yield an amplifier with the maximum possible power gain.

Cousequently, some form of device is needed to effectively "stretch" the output pulse from the first reactor so that it will match the wave form of the second. Various techniques have been tried to accomplish this.

Storage elements such as capacitors or inductors can be used to absorb the energy during the firing interval of the first reactor and deliver it a short time later to the output core, but at a much slower rate, so that the pulse will be more effective. Unfortunately, it is difficult to store the energy and in most practical cases much of it is delivered to the output core, not during its reset cycle but during its firing cycle, thereby wasting it. Consequently, the amplifier's figure of merit is substantially reduced.

Rather than devise lossless storage elements, the oversetting problem may be easily solved by opening the power winding with a mechanical switch during the reset interval. While this is technically possible and has been accomplished experimentally, it is not considered desirable for long life.

Diode ring modulators or transistors may be used to replace the mechanical switch but require several expensive components and have other undesirable characteristics.

Proposed Solution

Since all that is required is some means to block the signal in the power circuit during reset of the core, it may be observed by tracing out the polarities in Fig. 2, that this can be accomplished by inserting a counter voltage in the power circuit, as shown in Fig. 3.

This circuit is identical to Fig. 2 except that a half-wave pulsating signal has been placed in the power circuit to increase the blocking voltage applied to the power winding rectifier during core reset. With this blocking bias it is possible to increase the instantaneous rate of reset, i.e., if the blocking bias voltage is three times the normal power winding voltage, the core may be reset at a rate equal to four times normal. This means a limited range of control signal may be stretched, into one which is effectively much larger. Thus, with the power winding blocked, we are free to readjust the turns ratio of *SR* in Fig. 3 to obtain maximum energy transfer to the core. This permits much larger power gains per stage with no penalty in increased time response; hence, an increased amplifier figure of merit can be readily obtained.

There are numerous convenient methods of generating the signal for the stretch bias. One of the most convenient is to superimpose an a-c signal on a fully rectified d-c signal as shown in Fig. 4.

During the first half-cycle E_b subtracts from E_p reducing its value by one-third. During the next half-cycle E_b adds to the E_p thereby supplying twice the net blocking voltage to the rectifier.

With this arrangement E_p must be increased to compensate for voltage cancellation by E_b during the firing cycle. Since the power winding rectifier is now receiving the extra blocking voltage, the instantaneous control signal may be doubled without encountering overset difficulties. Fig. 5. shows a typical pulse stretch amplifier using this method.

Stretch may be applied up to five times the normal E_p value with favorable results. Increased stretch ratios seem to yield little improvement. Problems encountered are:

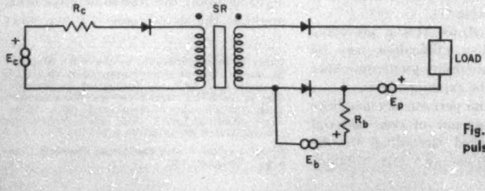


Fig. 3. Simplified pulse stretch amplifier circuit

1. The power winding rectifier leakage increases, resulting in uncontrolled core reset and loss of control range.

2. This leakage causes temperature drift problems.

3. The effective core B-H volume product increases. This is believed to be caused by an increase in the instantaneous value of core H caused by eddy current losses.

4. Energy lost in R_m increases.

These problems are all minimized by operating at stretch ratios of from 3 to 5.

Pulse stretch circuitry does not appear to be too practical for output stages. Iron volumes can usually be optimized, eliminating the need for pulse stretch compensation.

Conclusions

Pulse stretch circuits are most applicable to high-performance low-level stages which do not receive optimum input wave forms or where it is inconvenient to optimize core volumes.

The pulse stretch technique can be applied to many varieties of magnetic amplifiers. Its operation is simple, straightforward, and can be readily calculated.

It is hoped that pulse stretch circuits will afford another useful tool with which to attack difficult design problems.

References

1. ON THE CONTROL OF MAGNETIC AMPLIFIERS, R. A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 2124–28.

2. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, Robert A. Ramey. *Ibid.*, pp. 1214–23.

3. SINGLE CORE MAGNETIC AMPLIFIER AS A COMPUTING ELEMENT, R. A. Ramey. *Ibid.*, vol. 71, pt. I, 1952, pp. 442–46.

4. FAST-RESPONSE MAGNETIC AMPLIFIERS, Gerald E. Hughes, Harvey A. Miller. *Ibid.*, vol. 73, pt. I, Mar. 1954, pp. 69–75.

Temperature Stabilization of Transistor Amplifiers

L. M. VALLESE
ASSOCIATE MEMBER AIEE

Synopsis: A novel method of analysis of the thermal behavior of transistor amplifiers is developed by introducing the temperature incremental equations of the transistor and of its associated d-c network. Design formulas for prescribed thermal stability, and in particular for the case of thermistor compensation, are derived. Experimental verifications of the theory are included.

THE DESIGN of transistor amplifiers for thermal stability is usually performed on the basis of the equations expressing the derivatives (or the logarithmic derivatives) of collector current and voltage with respect to temperature. Shea[1] has taken into account the temperature dependence of I_{co} (open input-collector current), Tate[2] has added that of the collector resistance and, recently, Lin and Barco[3] have considered the temperature dependence of the input conductance of the transistor. In general, the expressions of the I_c and V_c derivatives obtained for certain fundamental types of d-c circuit configurations[4] are used directly for the thermal stability design. Temperature compensation can also be obtained with the use of thermistors and other temperature-dependent elements.

In the following, the problem of the temperature analysis of transistor amplifiers is approached with a different procedure; it is based on the observation that the rates of change of the input and output currents and voltages satisfy a set of linear equations obtained from the d-c characteristics of the transistor with an incremental procedure similar to that used to derive the so-called "small signal analysis" equations. The consideration of these "temperature incremental" equations provides an accurate and general method of thermal analysis and, in particular, permits the design of circuits containing thermistors. In the following, the simplifying assumption is made that all temperature rates of change are independent of temperature in the range of interest.

Temperature Incremental Equations

In general, if the transistor d-c characteristics are represented with the equations (referred to common base configuration)

$$I_e=f_1(V_{eb},\ V_{cb}) \qquad (1)$$
$$I_c=f_2(V_{eb},\ V_{cb})$$

differentiation with respect to T (absolute temperature) gives

$$\frac{dI_e}{dT}=\frac{\partial f_1}{\partial T}+\frac{\partial f_1}{\partial V_{eb}}\frac{dV_{eb}}{dT}+\frac{\partial f_1}{\partial V_{cb}}\frac{dV_{cb}}{dT} \qquad (2)$$
$$\frac{dI_c}{dT}=\frac{\partial f_2}{\partial T}+\frac{\partial f_2}{\partial V_{eb}}\frac{dV_{eb}}{dT}+\frac{\partial f_2}{\partial V_{cb}}\frac{dV_{cb}}{dT}$$

Since the coefficients $\partial f_1/\partial V_{eb}$, $\partial f_1/\partial V_{cb}$, $\partial f_2/\partial V_{eb}$, and $\partial f_2/\partial V_{cb}$ are computed at the quiescent point of operation, they are equal to the corresponding small signal analysis coefficients, when the transistor amplifier is operating in class A. In this case, using primes to indicate temperature derivatives and letting $I_{eT}' = \partial f_1/\partial T$, $I_{cT}' = \partial f_2/\partial T$, equations 2 are written

$$I_e'=I_{eT}'+g_{ee}V_{eb}'+g_{ec}V_{cb}' \qquad (3)$$
$$I_c'=I_{cT}'+g_{ce}V_{eb}'+g_{cc}V_{cb}'$$

When the transistor does not operate in class A, equations 2 should be used instead of 3. Equations 2 and 3 are the desired temperature incremental equations of the transistor; for class A operation their equivalent circuit representation shown in Fig. 1, case A, may be considered obtained from the small signal equivalent circuit with the introduction of current sources I_{eT}', I_{cT}'.

Shockley, et al.,[5] have analyzed the temperature dependence of the d-c transistor equations. Letting, for the common base configuration

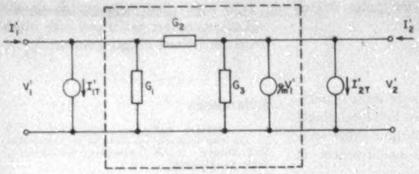
$$I_e=\frac{G_{ee}}{\Lambda}(e^{\Lambda V_{eb}}-1)+\frac{G_{ec}}{\Lambda}(e^{\Lambda V_{cb}}-1) \qquad (4)$$
$$I_c=\frac{G_{ce}}{\Lambda}(e^{\Lambda V_{eb}}-1)+\frac{G_{cc}}{\Lambda}(e^{\Lambda V_{cb}}-1)$$

where $\Lambda = +q/kT$ for p-n-p and $-q/kT$

Paper 56-276, recommended by the AIEE Solid State Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted November 30, 1955; made available for printing, March 26, 1956.

L. M. VALLESE is with Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

This work was performed at the Microwave Research Institute of the Polytechnic Institute of Brooklyn, under a Contract sponsored by the Office of Naval Research. The assistance of B.S. Fabricant in obtaining the experimental results of Figs. 3 and 5 is gladly acknowledged by the author.



Case A—$G_1 = g_{ee} + g_{ec}$, $G_2 = -g_{ec}$, $g_m = g_{ce} - g_{ec}$, $G_3 = g_{cc} + g_{ec}$

Case B—$G_1 = g_{bb} + g_{bc}$, $G_2 = -g_{bc}$, $g_m = g_{cb} - g_{bc}$, $G_3 = g_{cc} + g_{bc}$

Fig. 1. Temperature incremental equivalent circuit of transistor

for n-p-n transistors $G_{ce} = G_{ec}$, it is found that the G coefficients are proportional to σ_i^2/σ_n or σ_i^2/σ_n, where σ_i, σ_p, and σ_n are respectively the intrinsic, the p-type, and the n-type conductivities. Since

$$\sigma_i^2 \propto \exp(-E_g/kT)$$

where E_g is the energy gap, and σ_p, σ_n are proportional to $(T)^{-3/2}$ at room temperature, then

$$\frac{\partial}{\partial T}\left(\frac{G}{\Lambda}\right) \propto \frac{d}{dT}\left[T^{3/2}\exp(-E_g/kT)\right] = \left(\frac{5}{2T}+\frac{E_g}{kT^2}\right)\left[T^{3/2}\exp(-E_g/kT)\right] \quad (5)$$

In other words, at room temperature and with $E_g = 0.72eV$ for germanium, the G coefficients increase

$$\left(\frac{5}{2T}+\frac{E_g}{kT^2}\right)\% \cong 10\%$$

per K (degree Kelvin).

Accordingly, the terms I_{eT}', I_{cT}' in equations 3 become respectively

$$I_{eT}' \cong -0.1\left(\frac{G_{ee}+G_{ec}}{\Lambda}\right)+\left(\frac{0.1}{\Lambda}-\frac{V_{eb}}{T}\right)\times G_{ee}\epsilon^{\Lambda V_{eb}}+\left(\frac{0.1}{\Lambda}-\frac{V_{cb}}{T}\right)G_{ec}\epsilon^{\Lambda V_{cb}} \quad (6)$$

$$I_{cT}' \cong -0.1\left(\frac{G_{ce}+G_{cc}}{\Lambda}\right)+\left(\frac{0.1}{\Lambda}-\frac{V_{eb}}{T}\right)\times G_{ce}\epsilon^{\Lambda V_{eb}}+\left(\frac{0.1}{\Lambda}-\frac{V_{cb}}{T}\right)G_{cc}\epsilon^{\Lambda V_{cb}}$$

Usually these may be approximated as follows:

$$I_{eT}' \cong \left(\frac{0.1}{\Lambda}-\frac{V_{eb}}{T}\right)G_{ee}\epsilon^{\Lambda V_{eb}}$$

$$I_{cT}' \cong \left(\frac{0.1}{\Lambda}-\frac{V_{eb}}{T}\right)G_{ce}\epsilon^{\Lambda V_{eb}} \quad (7)$$

For example, using the values of d-c conductance coefficients published by Giacoletto[6] for a p-n-p alloy transistor at 298 K

$G_{ee} = 75.2$ micromhos, $G_{ec} = G_{ce} = -70.5$ micromhos, $G_{cc} = 137$ micromhos

$\Lambda = 38.8\ V^{-1}$, $V_{eb} = 0.17\ V$, $V_{cb} = -6V$, $I_e = 1$ milliampere

one finds

$$I_{eT}' \cong 111\times 10^{-6}\ A \text{ per K}$$

$$I_{cT}' \cong -105\times 10^{-6}\ A \text{ per K} \quad (8)$$

The temperature incremental equations for the common emitter configuration may be derived similarly from the corresponding d-c characteristics

$$I_b = \frac{1}{\Lambda}(G_{ee}+G_{ec}+G_{ce}+G_{cc}) - \frac{1}{\Lambda}\left[G_{ee}+G_{ec}+(G_{ec}+G_{cc})\epsilon^{\Lambda V_{ce}}\right]\epsilon^{-\Lambda V_{be}} \quad (9)$$

$$I_c = \frac{-1}{\Lambda}(G_{ce}+G_{cc})+\frac{1}{\Lambda}(G_{ce}+G_{cc}\epsilon^{\Lambda V_{ce}})\epsilon^{-\Lambda V_{be}}$$

Of course, the equations may be obtained more simply from equations 2 or 3, letting $I_{bT}' = -I_{eT}' - I_{cT}'$. For class-$A$ operation they are written

$$I_b' = I_{bT}' + g_{bb}V_{be}' + g_{bc}V_{ce}' \quad (10)$$

$$I_c' = I_{cT}' + g_{cb}V_{be}' + g_{cc}V_{ce}'$$

where

$$g_{bb} = g_{ee}+g_{ec}+g_{ce}+g_{cc},\quad g_{bc} = -g_{ec}-g_{cc},\quad g_{cb} = -g_{ce}-g_{cc},$$

and

$$I_{bT}' \cong \frac{0.1}{\Lambda}(G_{ee}+G_{ec}+G_{ce}+G_{cc}) - \left(\frac{0.1}{\Lambda}-\frac{V_{eb}}{T}\right)(G_{ee}+G_{ce})\epsilon^{\Lambda V_{eb}} - \left(\frac{0.1}{\Lambda}-\frac{V_{cb}}{T}\right)(G_{ec}+G_{cc})\epsilon^{\Lambda V_{cb}}$$

I_{cT}' has been given in equation 6. The equivalent circuit representation is shown in Fig. 1, case B. With the data of the example previously considered, one finds

$$I_{bT}' \cong -6.8\times 10^{-6}\ A \text{ per K} \quad (11)$$

$$I_{cT}' \cong -105\times 10^{-6}\ A \text{ per K}$$

In practice, the actual transistor characteristics differ from equations 4 because of the presence of extrinsic elements, such as base-spread and collector-leakage resistances. Accordingly, equations 4 should be modified to include these additional elements. For example, on account of the base spread resistance r_{bb}', the quantities V_{eb} and V_{cb} in equation 4 should be replaced with $V_{eb}+I_br_{bb}'$ and $V_{cb}+I_br_{bb}'$ respectively. Another less rigorous procedure consists in replacing equations 4 with experimentally determined characteristics, provided these may be expressed analytically with good approximation. Fig. 2 represents typical input and output characteristics of a p-n-p common emitter transistor. For the linear region of interest they may be represented approximately with the equations

$$I_b = g_{bb}(V_{be}-V_{be}^\circ)+g_{bc}(V_{ce}-V_{ce}^\circ) \quad (12)$$

$$I_c = I_{coe}+\alpha_{cb}I_b+g_{cc}^\circ(V_{ce}-V_{ce}^\circ)$$

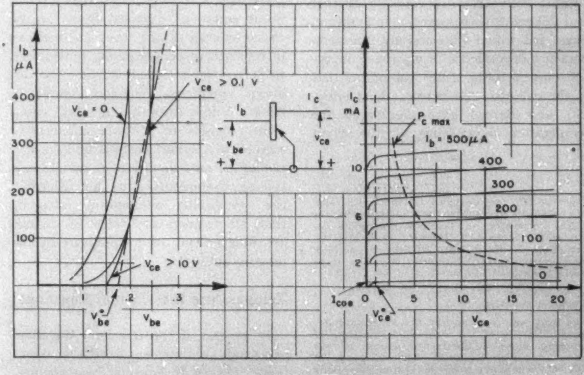


Fig. 2. Typical input and output characteristics of common emitter transistor

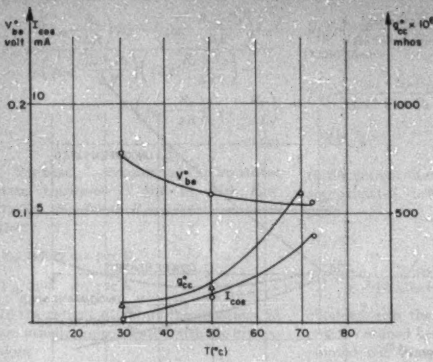


Fig. 3. Summary of thermal behavior of a common emitter transistor

where V_{be}°, V_{ce}° are shown in Fig. 2, and $I_{cos}=(1+\alpha_{cb})I_{cob}$ is the collector current for $I_b=0$, V_{ce}°. Note that g_{ce}° is the output conductance with open input circuit and is equal to $g_{ce}-\alpha_{cb}g_{bc}$. From equations 12 the following expressions of $I_{bT}'=\partial I_b/\partial T$, $I_{cT}'^*=\partial I_c/\partial T$ are derived

$$I_{bT}'=\frac{\partial g_{bb}}{\partial T}(V_{be}-V_{be}^\circ)-g_{bb}\frac{\partial V_{be}^\circ}{\partial T}+\frac{\partial g_{bc}}{\partial T}(V_{ce}-V_{ce}^\circ) \quad (13)$$

$$I_{cT}'^*=\frac{\partial}{\partial T}I_{coe}+\frac{\partial \alpha_{cb}}{\partial T}I_b+\frac{\partial g_{ce}}{\partial T}(V_{ce}-V_{ce}^\circ)$$

In first approximation these may be written

$$I_{bT}'\cong\frac{\partial g_{bb}}{\partial T}(V_{be}-V_{be}^\circ)-g_{bb}\frac{\partial V_{be}^\circ}{\partial T}$$
$$(14)$$
$$I_{cT}'^*\cong\frac{\partial}{\partial T}I_{coe}+\frac{\partial g_{ce}^\circ}{\partial T}(V_{ce}-V_{ce}^\circ)$$

It is noted that

$$I_{cT}'^*=\left(\frac{\partial I_c}{\partial T}\right)_{I_b,\,V_{ce}\,\mathrm{const}}$$
$$\neq I_{cT}'=\left(\frac{\partial I_c}{\partial T}\right)_{V_{be},\,V_{ce}\,\mathrm{const}}$$

In fact, $I_{cT}'=I_{cT}'^*+\alpha_{cb}I_{bT}'$.

The temperature incremental equations derived from equations 12 are

$$I_b'=I_{bT}'+g_{bb}V_{be}'+g_{bc}V_{ce}'$$
$$(15)$$
$$I_c'=I_{cT}'^*+\alpha_{cb}I_b'+g_{ce}^\circ V_{ce}'$$

They may be written as in equation 10 provided

$$I_{cT}'=I_{cT}'^*+\alpha_{cb}I_{bT}',\quad g_{cb}=\alpha_{cb}g_{bb}$$
$$(16)$$
$$g_{ce}=g_{ce}^\circ+\alpha_{cb}g_{bc}$$

In Fig. 3 are summarized the results of an experimental investigation of the thermal behavior of a common emitter p-n-p alloy transistor. With reference to the operating point $V_{be}=-0.2V$, $I_b=-60\mu A$, $V_{ce}=-6V$ at 30 degrees centigrade, one finds

$g_{bb}=1.35\times10^{-3}$ mhos,
$g_{bc}=-0.5\times10^{-6}$ mhos,
$g_{ce}^\circ=70$ micromhos

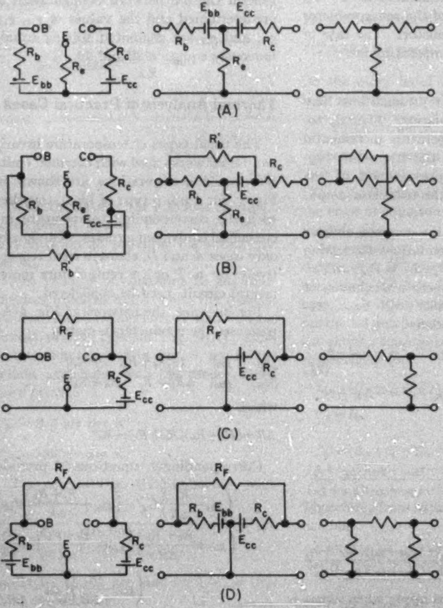


Fig. 4 (left). Temperature invariant d-c bias networks for class-A transistor amplifiers

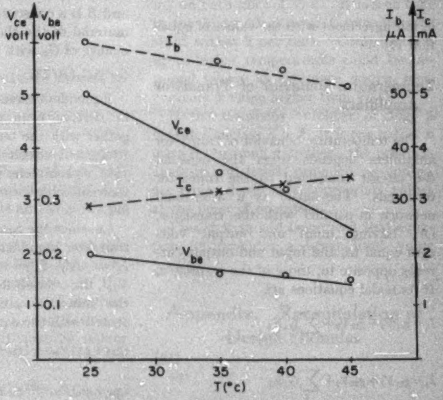


Fig. 5 (above). Thermal variations of V_{be}, V_{ce}, I_c, and I_b, in a common emitter-transistor amplifier

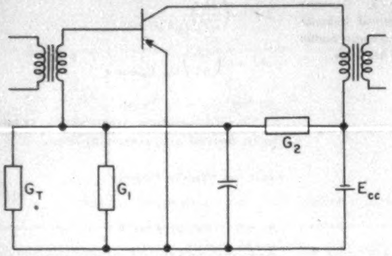


Fig. 6. Thermistor-compensated common emitter-transistor amplifier

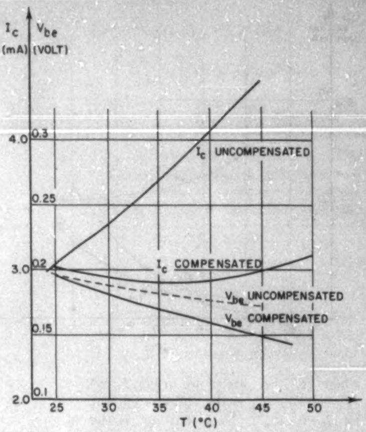


Fig. 7. Thermal behavior of the circuit of Fig. 6: (a) with thermistor, (b) with ohmic resistor replacing thermistor

$$\alpha_{cb}=40, \quad V_{be}{}^{\circ}=-0.155\ V,\ I_{coe}=-0.1 \text{ milliampere}$$

$$V_{ce}{}^{\circ}=-1\,V,\ \frac{\partial g_{bb}}{\partial T}\cong 0,$$

$$\frac{\partial V_{be}{}^{\circ}}{\partial T}\cong +0.0018\ V \text{ per K} \quad (17)$$

$$\frac{\partial I_{coe}}{\partial T}\cong 0.1 I_{coe},\ \frac{\partial g_{ec}{}^{\circ}}{\partial T}\cong 0.1\ g_{ec}{}^{\circ}$$

Therefore, from equations 14

$$I_{bT}'\cong -2.44\times 10^{-6}\ A \text{ per K}$$

$$I_{cT}'^{*}\cong -45\times 10^{-6}\ A \text{ per K} \quad (18)$$

$$I_{cT}'\cong -142.6\times 10^{-6}\ A \text{ per K}$$

in fair agreement with the values of equation 11.

Temperature Behavior of Transistor Amplifiers

The temperature behavior of transistor amplifiers depends upon the external d-c circuit connected to the transistor terminals. The latter is a 3-terminal network in parallel with the transistor, i.e., having input and output voltages equal to, and input and output currents opposite to, those of the transistor. If its nodal equations are

$$I_1=g_{11}V_1+g_{12}V_2+\sum_i G_iE_i$$
$$(19)$$
$$I_2=g_{21}V_1+g_{22}V_2+\sum_j G_jE_j$$

the corresponding temperature incremental equations are written

$$I_1'=g_{11}V_1'+g_{12}V_2'+\theta_1$$
$$(20)$$
$$I_2'=g_{21}V_1'+g_{22}V_2'+\theta_2$$

In equations 20 θ_1 and θ_2 are respectively the partial derivatives of I_1 and I_2 with respect to temperature. If the network is temperature invariant, $\theta_1=\theta_2=0$; if the network in parallel with the transistor, i.e., contains thermistors (in particular p-n junction diodes), θ_1 and θ_2 are computed on the basis of the temperature dependence of their conductances. This is approximately of the type[7]

$$G_T\cong G_{T_0}e^{B/T_0}e^{-B/T}$$

where G_{T_0} is the conductance at $T=T_0$, and B is a constant of the semiconductor material of the thermistor. The rate of change of G_T with temperature is

$$G_T'\cong G_T(B/T^2)$$

Equivalent circuit representations may be derived from equations 20 and, together with the temperature incremental equivalent circuit of the transistor, provide a complete representation of the thermal behavior of the transistor amplifier.

Consider the case of a common emitter transistor amplifier, equations 10, letting $I_1'=-I_b'$, $I_2'=-I_c'$, $V_1'=V_{be}'$, $V_2'=V_{ce}'$, the over-all temperature behavior of the transistor amplifier will be represented with the equations

$$(g_{bb}+g_{11})V_{be}'+(g_{bc}+g_{12})V_{ce}'=-I_{bT}'-\theta_1$$
$$(21)$$
$$(g_{cb}+g_{21})V_{be}'+(g_{cc}+g_{22})V_{ce}'=-I_{cT}'-\theta_2$$

Then

$$V_{be}'=\frac{(g_{bc}+g_{12})(I_{cT}'+\theta_2)-(g_{cc}+g_{22})(I_{bT}'+\theta_1)}{(g_{bb}+g_{11})(g_{cc}+g_{22})-(g_{bc}+g_{12})(g_{cb}+g_{21})}$$
$$(22)$$
$$V_{ce}'=\frac{(g_{cb}+g_{21})(I_{bT}'+\theta_1)-(g_{bb}+g_{11})(I_{cT}'+\theta_2)}{(g_{bb}+g_{11})(g_{cc}+g_{22})-(g_{bc}+g_{12})(g_{cb}+g_{21})}$$

The same equations apply when equations 15 instead of 10 are given, provided the relations 16 are used. From equations 22 and equations 10 or 20, the values of I_b', I_c' follow.

Equations 22 may be used for the design of temperature-stabilized amplifiers, with prescribed values of V_{be}', V_{ce}' (usually I_c' and V_{ce}' are prescribed). In general the d-c network configuration is predetermined and the values of g_{11}, g_{12}, g_{21}, and g_{22} are computed solving simultaneously equation 22 and 19.

Thermal Analysis of Practical Cases

The usual types of temperature-invariant d-c networks used with common emitter class-*A* configurations are shown in Fig. 4 (for a p-n-p type of bias), together with the corresponding temperature incremental equivalent circuits. Obviously only cases *A* and *D*, characterized respectively by a *T* or a π temperature incremental circuit, need be considered.

For instance, the network of Fig. 4(A) possesses the conductance matrix

$$\left\|\begin{matrix} g_{11} & g_{12} \\ g_{21} & g_{22}\end{matrix}\right\| = \frac{1}{\Delta R}\left\|\begin{matrix} R_e+R_c, & -R_e \\ -R_e & R_e+R_b\end{matrix}\right\|$$

where

$$\Delta R=(R_e+R_b)(R_e+R_c)-R_e^2$$

Correspondingly, equations 22 provide

$$V_{be}'=\frac{\left(g_{bc}-\frac{R_e}{\Delta R}\right)I_{cT}'-\left(g_{cc}+\frac{R_e+R_b}{\Delta R}\right)I_{bT}'}{\left(g_{bb}+\frac{R_e+R_c}{\Delta R}\right)\left(g_{cc}+\frac{R_e+R_b}{\Delta R}\right)-\left(g_{bc}-\frac{R_e}{\Delta R}\right)\left(g_{cb}-\frac{R_e}{\Delta R}\right)}$$

$$V_{ce}'=\frac{\left(g_{cb}-\frac{R_e}{\Delta R}\right)I_{bT}'-\left(g_{bb}+\frac{R_e+R_b}{\Delta R}\right)I_{cT}'}{\left(g_{bb}+\frac{R_e+R_c}{\Delta R}\right)\left(g_{cc}+\frac{R_e+R_b}{\Delta R}\right)-\left(g_{bc}-\frac{R_e}{\Delta R}\right)\left(g_{cb}-\frac{R_e}{\Delta R}\right)} \tag{23}$$

For example, considering the transistor data, equations 17 and 18, with $R_b=3{,}000\Omega$, $R_e=200\Omega$, $R_c=5{,}000\Omega$, one finds that

$V_{be}'\cong 2.22$ mv per K

$V_{ce}'\cong 95$ mv per K

In the current literature the equations 23 are usually approximated with the expressions

$$V_{be}'\cong 0$$

$$V_{ce}'\cong\frac{-I_{coe}'}{g_{ce}^{\circ}+\alpha_{cb}\frac{R_e}{\Delta R}+\frac{R_e+R_b}{\Delta R}} \tag{24}$$

obtained, assuming $I_{bT}'=0$, $V_{be}'=0$, $I_{cT}'=I_{coe}'$ and substituting I_b' of equations 15 with the value provided by the first of equations 20. Note that, due to the approximations involved, the result equation 24 does not represent a unique solution; as a matter of fact from equation 23 with the same assumptions one obtains the following:

$$V_{ce}'=\frac{-I_{coe}'}{g_{ce}^{\circ}+\alpha_{cb}g_{bc}+\frac{R_e+R_b}{\Delta R}-\frac{\left(g_{bc}-\frac{R_e}{\Delta R}\right)\left(g_{cb}-\frac{R_e}{\Delta R}\right)}{g_{bb}+\frac{R_e+R_c}{\Delta R}}} \tag{25}$$

With the data of the previous example, equations 24 provide $V_{ce}'=13$ mv per K and equation 25 provides $V_{ce}'=15.2$ mv per K.

The validity of equations 23 has been checked experimentally using the transistor described in equations 17 and 18 in a circuit of the type in Fig. 4(A). The results, summarized in Fig. 5, are in excellent agreement with the theory, since they provide

$V_{be}'=2.5$ mv per K

$V_{ce}'=132$ mv per K

The case of Fig. 4(D) is solved, with equations

$$\left\|\begin{matrix} g_{11} & g_{12} \\ g_{21} & g_{22}\end{matrix}\right\|=\left\|\begin{matrix} G_b+G_P, & -G_P \\ -G_P & G_c+G_P\end{matrix}\right\|$$

where G_b, G_P, G_c are used for $1/R_b$, $1/R_P$, $1/R_c$ respectively.

$$V_{be}'=\frac{(g_{bc}-G_P)I_{cT}'-(g_{cc}+G_c+G_P)I_{bT}'}{(g_{bb}+G_b+G_P)(g_{cc}+G_c+G_P)-(g_{bc}-G_P)(g_{cb}-G_P)}$$

$$V_{ce}'=\frac{(g_{cb}-G_P)I_{bT}'-(g_{bb}+G_b+G_P)I_{cT}'}{(g_{bb}+G_b+G_P)(g_{cc}+G_c+G_P)-(g_{bc}-G_P)(g_{cb}-G_P)} \tag{26}$$

In the current literature equations 26 are approximated, usually with the expressions

$$V_{be}'\cong 0$$

$$V_{ce}'\cong\frac{-I_{coe}'}{g_{ce}^{\circ}+\alpha_{cb}G_P+G_c+G_P} \tag{27}$$

obtained with the assumptions and procedure outlined before. Again it may be pointed out that, due to the approximations involved, equations 27 do not represent a unique solution, since from equations 26 with the same assumptions one obtains

$$V_{ce}'=\frac{-I_{coe}'}{g_{cc}^{\circ}+\alpha_{cb}g_{bc}+G_c+G_P-\frac{(g_{bc}-G_P)(g_{cb}-G_P)}{g_{bb}+G_b+G_P}} \tag{28}$$

For example, with the data of equations 17 and 18 with $R_b=\infty$, $R_P=60k\Omega$, $R_c=5{,}000\Omega$, one finds from equations 26

$V_{be}'\cong 2.38$ mv per K

$V_{ce}'\cong 46.5$ mv per K

On the other hand equations 27 and 28 both provide approximately

$V_{ce}'\cong 10.5$ mv per K

From the previous discussion it appears that in general I_{bT}', V_{be}', and $\partial g_{cc}^{\circ}/\partial T\,(V_{ce}-V_{ce}^{\circ})$ are not negligible, and that the error of equations 24 and 27 is not on the safe side.

An example of temperature variable d-c network is shown in Fig. 6 for a common emitter class-*A* amplifier. In this case, if G_T is the conductance of the thermistor (at temperature *T*) and G_1 that of its shunt, equations 19 and 20 reduce respectively to

$$\begin{aligned} -I_b &=(G_T+G_1+G_2)V_{be}-G_2E_{cc} \\ V_{ce}&=E_{cc}\end{aligned} \tag{19A}$$

$$\begin{aligned} -I_b' &=(G_T+G_1+G_2)V_{be}'+G_T'V_{be} \\ V_{ce}'&=0\end{aligned} \tag{20A}$$

Therefore, from equations 20(A) and 10 one has

$$V_{be}'=\frac{-I_{bT}'-G_T'V_{be}}{g_{bb}+G_T+G_1+G_2} \tag{29}$$

$$I_c'=I_{cT}'+g_{cb}V_{be}' \tag{30}$$

As an example, consider again the transistor data, equations 17 and 18, and assume that the circuit of Fig. 6 must be designed for $I_c'=0$. A thermistor, which presents a conductance $G_T=2.7\times10^{-4}$ mhos (3,700Ω) at room temperature under an applied difference of potential $=0.2$ volt, is available. Its *B* constant is 3,500 K, its G_T' is $\sim1.05\times10^{-6}$ mhos per K. From equations 29 and 30 follows

$$I_{cT}'(g_{bb}+G_T+G_1+G_2)=g_{cb}(I_{bT}'+G_T'V_{be}) \tag{31}$$

On the other hand, substituting in equation 31, the relation

$$G_T+G_1+G_2=\frac{G_2E_{cc}-I_b}{V_{be}}$$

obtained from equation 19(A) gives the equation

$$G_2=\left(g_{cb}\frac{I_{bT}'+G_T'V_{be}}{I_{cT}'}-g_{bb}\right)\frac{V_{be}}{E_{cc}}+\frac{I_b}{E_{cc}} \tag{32}$$

With the values of the example considered it is found that $G_2=22.4\times10^{-6}$ mhos (44,700Ω) and $G_1=78\times10^{-6}$ mhos (12,800Ω).

In Fig. 7 the results of the experimental verification are shown. Having taken as reference the values at 25 C (degrees centigrade), the collector current is compensated in a temperature range extending on both sides of 25 C. It is seen that in the 25 to 50 interval I_c remains constant within 2 per cent. Compensation up to higher temperatures could be obtained taking as reference design temperature a value higher than 25 C.

If the transistor amplifier of Fig. 6 operates in class *B* (e.g., it is one half of a class-*B* push-pull amplifier), the previous analysis still holds, provided g_{bb} and g_{cb} are replaced with the appropriate values at the point of operation.

Finally it should be noticed that equation 32 provides a lower limit value for G_T.

Appendix. Recapitulation of Design Formulas

Let

$$I_{bT}'\cong\frac{\partial g_{bb}}{\partial T}(V_{be}-V_{be}^{\circ})-g_{bb}\frac{\partial V_{be}^{\circ}}{\partial T}\quad(\cong-2\times10^{-3}g_{bb}\ A\text{ per }K)$$

$$\begin{aligned} I_{cT}'&\cong\frac{\partial I_{coe}}{\partial T}+\frac{\partial\alpha_{cb}}{\partial T}+\frac{\partial g_{cc}^{\circ}}{\partial T}\times(V_{ce}-V_{ce}^{\circ})+\alpha_{cb}I_{bT}' \\ &\cong 0.1[I_{coe}+g_{cc}^{\circ}(V_{ce}-V_{ce}^{\circ})]+\alpha_{cb}I_{bT}'\end{aligned}$$

$$\Delta_g=(g_{bb}+g_{11})(g_{cc}+g_{22})-(g_{bc}+g_{12})(g_{cb}+g_{21})$$

where g_{bc} may be neglected in general.

The fundamental design equations for the temperature stabilization of a common emitter transistor amplifier are

$$V_{be}'\cong\frac{1}{\Delta_g}[(g_{bc}+g_{12})(I_{cT}'+\theta_2)-(g_{cc}+g_{22})\times (I_{bT}'+\theta_1)]$$

$$V_{ce}'\cong\frac{1}{\Delta_g}[(g_{cb}+g_{21})(I_{bT}'+\theta_1)-(g_{bb}+g_{11})\times (I_{cT}'+\theta_2)]$$

$$I_b'\cong I_{bT}'+g_{bb}V_{be}'+g_{bc}V_{ce}'=$$

$$\cong I_{bT}'+\frac{g_{bb}}{\Delta_g}[g_{12}(I_{cT}'+\theta_2)-(g_{cc}+g_{22})(I_{bT}'+\theta_1)]$$

$$I_c' \cong I_{cT}'+g_{cb}V_{be}'+g_{cc}V_{ce}'=$$

$$\cong\frac{1}{\Delta_g}\{[g_{cb}(g_{bc}+g_{12})-g_{cc}(g_{bb}+g_{11})]\times (I_{cT}'+\theta_2)+[g_{cc}(g_{cb}+g_{21})-g_{cb}\times (g_{cc}+g_{22})](I_{bT}'+\theta_1)\}+I_{cT}'$$

In the case of temperature invariant d-c bias networks, $\theta_1=\theta_2=0$. Correspondingly, the equation for I_c' may be simplified as follows:

$$I_c'\cong\frac{1}{\Delta_g}\{[g_{22}(g_{bb}+g_{11})-g_{21}(g_{bc}+g_{12})]I_{cT}'+(g_{cc}g_{21}-g_{cb}g_{22})I_{bT}'\}$$

Finally, in the case of temperature dependent d-c bias networks, if in particular V_{ce}=constant, i.e., $V_{ce}'=0$, the design equations are simplified as follows:

$$V_{be}'\cong\frac{-(I_{bT}'+\theta_1)}{g_{bb}+g_{11}}$$

$$V_{ce}'=0$$

$$I_b' \cong\frac{g_{11}I_{bT}'-g_{bb}\theta_1}{g_{bb}+g_{11}}$$

$$I_c' \cong I_{cT}'-\frac{g_{cb}}{g_{bb}+g_{11}}(I_{bT}'+\theta_1)$$

References

1. Transistor Operation: Stabilization of Operating Points, **R. F. Shea.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, Nov. 1952, pp. 1435–37.

2. Temperature Stabilized Transistor Amplifiers, **H. J. Tate.** *Electronics*, New York, N. Y., June 1954, p. 144.

3. Temperature Effects in Circuits Using Junction Transistors, **H. C. Lin, A. A. Barco.** *Transistors I*, Radio Corporation of America, Princeton, N. J., 1956, pp. 369–402.

4. Transistor Electronics (book), **A. W. Lo, R. O. Endres, J. Zawels, F. D. Waldhauer, C. C. Cheng.** Prentice-Hall, Inc., New ,York, N. Y., 1955, chap. 4.

5. P-N Junction Transistors, **W. Shockley, M. Sparks, G. K. Teal.** *The Physical Review*, New York, N. Y., vol. 83, July 1951, pp. 151–62.

6. Study of PNP Alloy Junction Transistor from D-C Through Medium Frequencies, **L. J. Giacoletto.** *RCA Review*, Radio Corporation of America, Princeton, N. J., vol. XV, no. 4, Dec. 1954, pp. 506–62.

7. Properties and Uses of Thermistors, **J. A. Becker, C. B. Green, G. L. Pearson.** *AIEE Transactions*, vol. 65, Nov. 1946, pp. 711–25.

Magnetic Frequency Multipliers and Their Rating

Part I—Frequency Triplers

WILLIAM McMURRAY
ASSOCIATE MEMBER AIEE

THERE ARE a growing number of applications in which a source of power at a higher frequency than the standard 60 or 400 cps (cycles per second) is either necessary or desirable. Rotating machines, electronic generators, and magnetic frequency multipliers are some of the devices which may be employed to obtain such a power supply. Magnetic frequency multipliers have certain inherent advantages which are leading to an expansion of their use. In this respect they are quite similar to magnetic amplifiers. Both devices consist of rugged, static components which have a high degree of reliability with a long life and which require little or no maintenance. They are available for use immediately after power is applied. They can be designed to meet the requirements of military specifications for operation under vibration, high shock, and extremes in ambient temperature. The major undesirable features of magnetic frequency multipliers are the low-input power factor and the relatively large size of the equipment. With the development of better core materials and circuits, a considerable improvement in these features has recently been achieved. Power at frequencies of two, three, four, five, or six times the input frequency may be generated with reasonable efficiency.

In any particular application, the use of a magnetic frequency multiplier is determined by the relative importance of its advantages and disadvantages in that application. Factors which must be considered are system performance, size, cost, efficiency, reliability, and maintenance required. Where a high-frequency power supply is necessary to obtain satisfactory performance of a system, the static frequency multiplier should be compared with other types of frequency converters of the same rating. In other applications, a high-frequency source may be a desirable alternative to the available power supply. Here the relative merit of the over-all system with and without a frequency converter should be considered.

The object of this paper is to present a method, with examples, for analyzing the performance and size of magnetic frequency multipliers. Part I deals with frequency triplers; Part II, dealing with frequency sextuplers, is planned for future presentation.

Nomenclature

e = voltage, instantaneous, volts
E = voltage, rms value, volts
E_m = maximum value of line-to-neutral voltage, volts (base of per-unit voltage)
f = frequency of a-c supply, cps
F = rating factor of a reactor
H = magnetomotive force, oersteds
i = current, instantaneous, amperes
I = current rms value, amperes
I_d = d-c bias, amperes (base of per-unit current)
N = number of turns in a reactor winding
pf = power factor
P = power, watts
R = load resistance, ohms
t = time, seconds
ϕ = magnetic flux, instantaneous, maxwells
Φ_s = magnetic flux at saturation, maxwells (base of per-unit flux)
$\omega=2\pi f$

Subscripts

a = average value
A,B,C = phases of 3-phase supply
c = choke
i = input winding
l = load
m = maximum value of a sinusoidal wave
N = star point
$n=2,3,4,5,\ldots$ = order number of harmonic
O = system neutral
o = output winding
p = peak value of a nonsinusoidal wave

In Figs. 3, 4, and 7 through 11, showing ideal voltage, flux, and current wave forms, the scales are marked in per-unit quantities, and unity turns ratio on all reactors has been assumed.

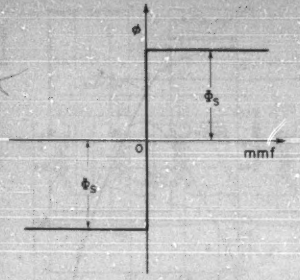


Fig. 1. Flux-magnetomotive force characteristic for ideal magnetic core material

Rating of Magnetic Frequency Multipliers

To estimate the size and cost of a magnetic frequency multiplier, it is necessary to know the volt-ampere ratings of the individual saturable reactors used in the circuit. Since these reactors are transformerlike devices, the expression of their rating in terms of an equivalent transformer rating is desirable, as it enables engineers familiar with transformer design and application to visualize readily the size of equipment required in a frequency multiplier. A similar method has been employed for many years in estimating the size of transformers for rectifiers.[1]

For a reactor used in a frequency multiplier, a rating factor F may be defined as the kva rating at the input frequency of a single-phase 2-winding transformer using an equivalent quantity of material, per kilovolt-ampere output of the frequency multiplier. Since both windings of the transformer would carry its kilovolt-ampere rating, the rating factor is calculated from the following expression, assuming resistive load

$$F=\frac{1}{2}\,\frac{\text{sum of volt-amperes of all windings}}{\text{maximum power output of complete circuit with resistive load}}$$

This factor will depend somewhat upon whether or not the multiplier output is filtered. The total rating factor of a frequency multiplier (i.e., the sum of the

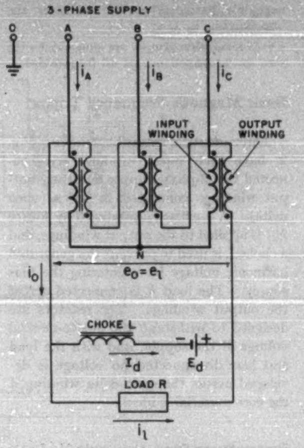


Fig. 2. Circuit diagram of basic magnetic frequency tripler

rating factors of the component reactors) is the total equivalent transformer kilovolt-ampere at supply frequency required to produce 1 kva of harmonic power.

Values of the rating factor are derived in this paper for two frequency tripler circuits employing d-c biased core reactors, as described qualitatively in reference 2. Such circuits have the advantage over circuits relying on a-c saturation[3,4] or ferroresonance[5] of being able to operate over a wide range of supply voltage and frequency variation, with only a directly proportional variation in output voltage and no change in wave shape. Furthermore, the output voltage and its wave shape do not change with load, apart from the internal impedance drop.

Also to be considered in determining the over-all size and cost of frequency multipliers are the size and cost of the d-c bias supply (usually a transformer and metallic rectifier), the bias circuit choke, the output transformer (if any), and the filter where used. The choke is about the same size as one of the multiplier reactors. The bias transformer is of much smaller rating. However, for units of low output rating these accessories form a substantial portion of the total cost.

Several differences between frequency multiplier reactors and ordinary transformers should be noted. The former are subjected to saturation and more rapid flux reversals, so that thinner core laminations may be required to reduce the eddy current loss. Also, rectangular loop core material is highly desirable. Hence the cores for frequency multipliers tend to be more expensive than transformer cores. However, to take advantage of the maximum difference between the unsaturated and saturated slopes of the magnetization curve, the cores may be designed for a greater flux swing than are transformer cores normally. Careful attention should be given during design to minimize leakage and after-saturation inductance. Furthermore, since there is relatively little

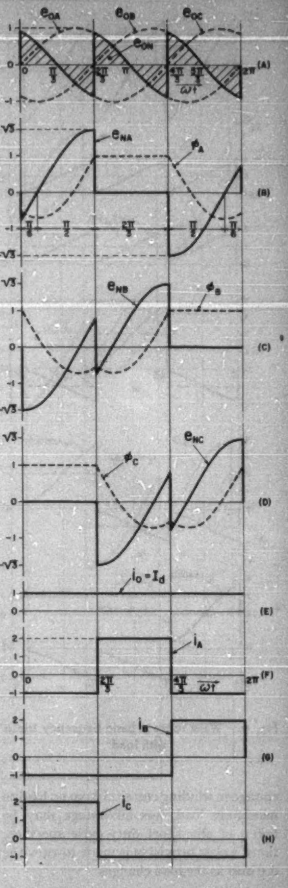


Fig. 3. Voltage, flux, and no-load current wave forms in basic frequency tripler

---

Paper 56-229, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted November 9, 1955; made available for printing April 10, 1956.

William McMurray is with the General Electric Company, Schenectady, N. Y.

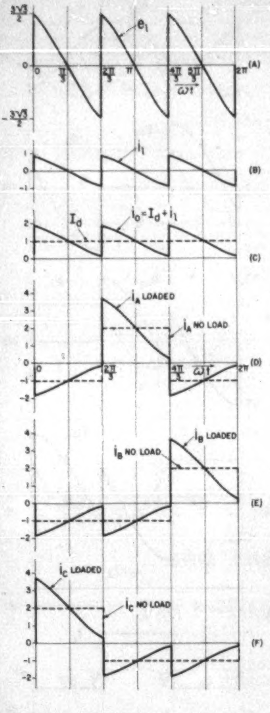


Fig. 4. Wave forms in basic frequency tripler with load

change in winding currents from no load to maximum load, less advantage may be taken of any short duty-cycle specifications, unless provision is made to vary the d-c bias as the load changes.

ASSUMPTIONS MADE IN ANALYSIS

In deriving the mode of operation of the magnetic frequency multiplier circuits and calculating the voltages and currents produced, the following assumptions are made to simplify the analysis.

1. The magnetic core material has a flux-magnetomotive-force relationship similar to that shown in Fig. 1. When the core is unsaturated, the net magnetomotive force acting upon it is negligible. After saturation, zero effective impedance is presented to currents flowing in the windings. Hysteresis and eddy currents losses are neglected.

2. Winding resistance and leakage inductance are neglected.

3. The auxiliary chokes have very large inductance and do not saturate, allowing negligible harmonic current to enter the bias supply.

4. The supply voltages are sinusoidal with zero source impedance at all frequencies.

Basic Magnetic Frequency Tripler

Three identical saturable reactors, Fig. 2, have their input windings wye-connected to a 3-phase supply and their output windings connected in series (open delta). A d-c bias I_d, produced by source E_d, is applied to the output windings, and a choke L is used to prevent the generated harmonic voltage from entering the bias supply. The load R is connected across the output windings. The reactors are designed to withstand the line-to-neutral voltage of the supply, and with the load and bias disconnected no voltage is developed across the open-delta winding if the core material is ideal.

STEADY-STATE OPERATION WITH NO LOAD

With direct current flowing in the output windings in a direction to bias the cores with conventionally positive polarity, Fig. 2, the wave forms of Fig. 3 are generated when steady-state conditions have been reached, as may be deduced from the following considerations.

At any given instant, at least one of the 3-phase input currents will flow toward the star point N of the input windings, aiding the magnetomotive force of the bias current. Therefore at least one reactor is saturated and unable to withstand voltage. However, at least two reactors must be unsaturated to withstand the 3-phase supply voltages. Hence at any time, one and only one of the reactors is saturated. Since the circuit is symmetrical with respect to the three phases, each reactor is saturated for one-third of every cycle. Also, the voltages existing between any two points of the input circuit can contain only alternating components. These conditions are satisfied only for the mode of operation illustrated in Fig. 3, in which saturation occurs from $2\pi/3$ to $4\pi/3$ radians of each phase voltage cycle. N follows the voltage of the line to which the instantaneously saturated reactor is connected. Hence the voltage e_{ON} between the system neutral O and N has the triple-frequency saw-toothed wave form of Fig. 3(A).

The individual reactor voltages and fluxes, Fig. 3(B), (C), and (D) are

$$e_{NA} = e_{OA} - e_{ON} \qquad (1)$$

$$e_{NB} = e_{OB} - e_{ON} \qquad (2)$$

$$e_{NC} = e_{OC} - e_{ON} \qquad (3)$$

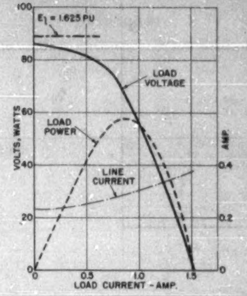


Fig. 5. Performance characteristics of basic frequency tripler

$$\phi_A = \frac{10^8}{N_i} \int e_{NA}\, dt \qquad (4)$$

The two reactors which are unsaturated at a given instant have zero net ampere-turns acting upon them; see Fig. 1. Hence a current equal to the bias current times N_i/N_o must flow out of each of their input windings, and a current of twice this value must flow into the third reactor, which is saturated. The total magnetomotive force, acting to saturate the third reactor, is therefore three times the bias ampere-turns. The input currents have the asymmetrical wave shapes of Fig. 3(F), (G), and (H).

The load voltage e_l, Fig. 4(A), is the sum of the voltages induced in the three series output windings e_o

$$e_l = e_O = -\frac{N_o}{N_i}(e_{NA} + e_{NB} + e_{NC})$$
$$= \frac{N_o}{N_i}(-e_{OA} + e_{ON} - e_{OB} + e_{ON} - e_{OC} + e_{ON}) \qquad (5)$$
$$= \frac{N_o}{N_i} 3e_{ON}$$

By Fourier analysis, this voltage may be expressed by the series

$$e_l = \frac{N_o}{N_i} E_m \frac{9\sqrt{3}}{\pi} \sum_{n=1}^{\infty} \frac{3_n}{(3n)^2 - 1} \sin 3n\omega t$$
$$= \frac{N_o}{N_i} E_m (1.861 \sin 3\omega t + 0.851 \sin 6\omega t \qquad (6)$$
$$+ 0.558 \sin 9\,\omega t + 0.416 \sin 12\omega t \ldots)$$

LOAD CONDITION

When the output windings are resistively loaded, a load current i_l of the same wave shape as the load voltage flows, Fig. 4(B), and is superposed upon the bias current in the output windings, Fig. 4(C). It is reflected into the input

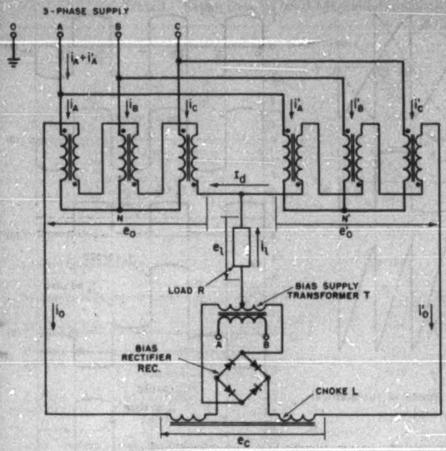


Fig. 6. Circuit diagram of symmetrical magnetic frequency tripler

currents as shown in Fig. 4(D), (E), (F). As the load resistance is decreased, a point will be reached at which the peak value of the load current equals the bias current. The corresponding value of the load resistance may be defined as the critical load resistance. For lower values of load resistance, the mode of operation of the circuit changes and the load voltage drops rapidly; see Fig. 5. This mode is not of practical interest and will not be discussed further. A detailed analysis is given in reference 6.

CALCULATION OF PERFORMANCE AND RATING OF BASIC FREQUENCY TRIPLER

The calculation of the performance of this circuit will be described in some detail, to illustrate the method which may be applied to most frequency multipliers using biased-core reactors. Unity turns ratio is assumed, to simplify the equations.

The load voltage, Fig. 4(A), is characterized by

peak value

$$E_{lP}=\frac{3\sqrt{3}}{2}E_m=2.598E_m \tag{7}$$

average value

$$E_{la}=\frac{3}{\pi}\int_0^{\frac{\pi}{3}}3E_m\sin\theta d\theta=\frac{9}{2\pi}E_m=1.432E_m \tag{8}$$

effective value

$$E_l=\sqrt{\frac{3}{\pi}\int_0^{\frac{\pi}{3}}(3E_m\sin\theta)^2d\theta}$$

$$=\frac{3}{\sqrt{2}}\sqrt{1-\frac{3\sqrt{3}}{4\pi}}E_m=1.625E_m \tag{9}$$

The voltage rating of each of the reactor windings at the input frequency is equal to the line-to-neutral voltage of the supply

$$E=\frac{E_m}{\sqrt{2}}=0.707\ E_m \tag{10}$$

For maximum power output into the load, the peak value of the load current I_{lp} is equal to the d-c bias I_d, and the critical load resistance is

$$R=\frac{E_{lp}}{I_d}=2.598\frac{E_m}{I_d} \tag{11}$$

Then the average and effective values of the maximum load current are

$$I_{la}=\frac{1.432}{2.598}I_d=0.511\ I_d \tag{12}$$

$$I_l=\frac{1.625}{2.598}I_d=0.625\ I_d \tag{13}$$

The maximum load power is therefore

$$P_{lp}=E_lI_l=1.625E_m\times0.625\ I_d$$
$$=1.016\ E_mI_d \tag{14}$$

At no load, the current in the output winding is the d-c bias alone

$$I_o=I_d \tag{15}$$

and the effective value of the current in the input winding is

$$I_i=\sqrt{\frac{2\ I_o^2+(2I_o)^2}{3}}=\sqrt{2}I_o \tag{16}$$
$$=1.414\ I_d$$

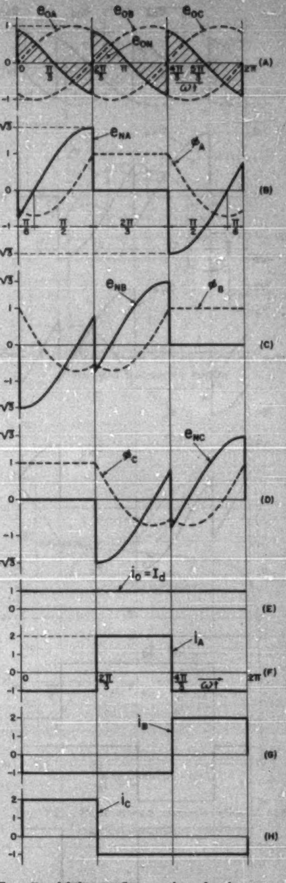


Fig. 7. Voltage, flux, and no-load current wave forms in first section of symmetrical frequency tripler

At maximum load, effective value of the current in the output winding is

$$I_o=\sqrt{I_d^2+(0.625\ I_d)^2}=1.179\ I_d \tag{17}$$

and the effective value of the current in the input winding is

$$I_i=\sqrt{2}I_o=1.667\ I_d \tag{18}$$

Thus the rating factor of one reactor is

$$F=\frac{0.707E_m(1.179I_d+1.667\ I_d)}{2\times1.016\ E_mI_d} \tag{19}$$
$$=0.990$$

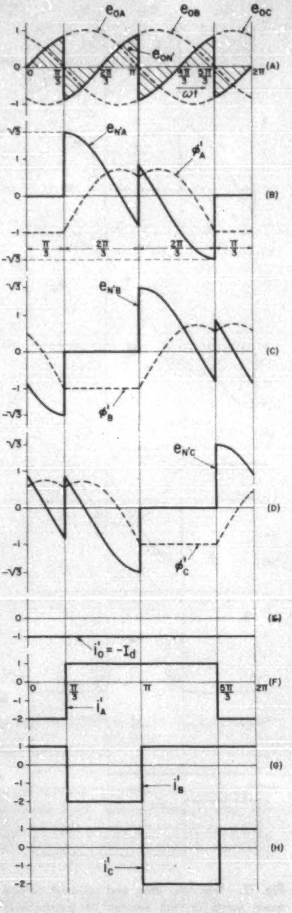


Fig. 8. Voltage, flux, and no-load current wave forms in second section of symmetrical frequency tripler

The total rating factor of the basic frequency tripler circuit is three times the value of equation 19

$$\sum F = 3F = 2.970 \tag{20}$$

Defining the effective power factor of the circuit as

$$\text{pf} = \frac{\text{power output into load}}{3 \times \text{rms phase voltage} \times \text{rms line current}} \tag{21}$$

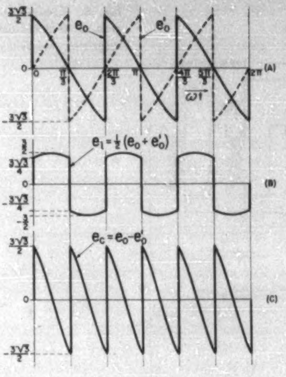


Fig. 9. Voltage wave forms in symmetrical frequency tripler

the maximum value for the power factor of the basic tripler is

$$\text{pf} = \frac{1.016\ E_m I_d}{3 \times 0.707\ E_m \times 1.667\ I_d} = 0.287 \tag{22}$$

These results have been summarized in Table I.

If an ideal filter were connected in series with the load, so that a pure third harmonic current is drawn, the calculations should be modified. The peak value of the load current, corresponding to maximum load power, is still equal to the d-c bias

$$I_{lm3} = I_d \tag{23}$$

The maximum third harmonic load power is therefore

$$P_{lm3} = \frac{1}{2} E_{lm3} \times I_{lm3} \tag{24}$$

$$= \frac{1}{2} \times 1.861 E_m \times I_d = 0.930 E_m I_d$$

At maximum load, the effective value of the current in the output winding is

$$I_o = \sqrt{I_d^2 + \frac{1}{2} I_d^2} = \sqrt{\frac{3}{2}} I_d = 1.225 I_d \tag{25}$$

The effective value of the current in the input winding is

$$I_1 = \sqrt{2} I_o = \sqrt{3} I_d = 1.732 I_d \tag{26}$$

Hence the rating factor of one reactor is

$$F = \frac{0.707\ E_m (1.225\ I_d + 1.732\ I_d)}{2 \times 0.930\ E_m I_d} = 1.124 \tag{27}$$

The total rating factor of the basic

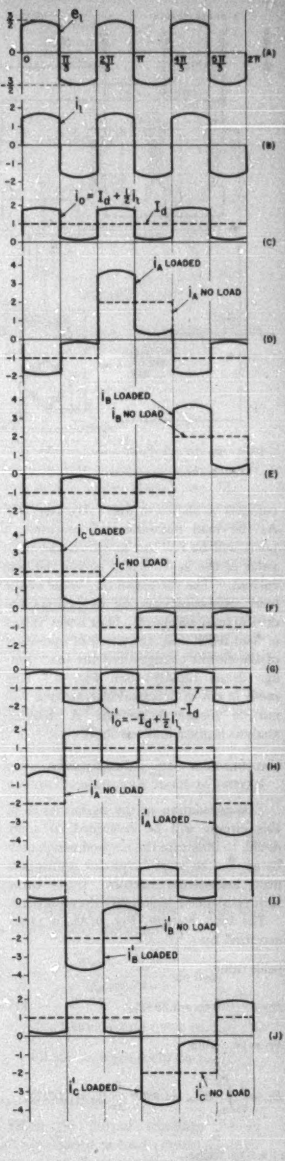


Fig. 10. Wave forms in symmetrical frequency tripler with load

Table I. Rating Data for Ideal Magnetic Frequency Triplers

		Basic Tripler		Symmetrical Tripler	
Load voltage	peak value	$\frac{3\sqrt{3}}{2}$	2.598	$\frac{3}{2}$	1.5
	average value	$\frac{9}{2\pi}$	1.432	$\frac{9}{2\pi}$	1.432
	effective rms value	$\frac{3}{\sqrt{2}}\sqrt{1-\frac{3\sqrt{3}}{4\pi}}$	1.625	$\frac{3}{2\sqrt{2}}\sqrt{1+\frac{3\sqrt{3}}{2\pi}}$	1.434
	maximum value of fundamental component	$\frac{27\sqrt{3}}{8\pi}$	1.861	$\frac{27\sqrt{3}}{8\pi}$	1.861
Voltage rating of reactor		$\frac{1}{\sqrt{2}}$	0.707	$\frac{1}{\sqrt{2}}$	0.707
Critical load resistance		$\frac{3\sqrt{3}}{2}$	2.598	$\frac{3}{4}$	0.75
Maximum load current, rms value		$\sqrt{\frac{2}{3}}\sqrt{1-\frac{3\sqrt{3}}{4\pi}}$	0.625	$\sqrt{2}\sqrt{1+\frac{3\sqrt{3}}{2\pi}}$	1.912
Maximum load power		$\sqrt{3}\left(1-\frac{3\sqrt{3}}{4\pi}\right)$	1.016	$\frac{3}{2}\left(1+\frac{3\sqrt{3}}{2\pi}\right)$	2.755
Current in output winding, rms value	no load	1	1.000	1	1.000
	maximum load	$\sqrt{\frac{5}{3}-\frac{\sqrt{3}}{2\pi}}$	1.179	$\sqrt{\frac{3}{2}}\sqrt{1+\frac{\sqrt{3}}{2\pi}}$	1.383
Current in input winding, rms value	no load	$\sqrt{2}$	1.414	$\sqrt{2}$	1.414
	maximum load	$\sqrt{2}\sqrt{\frac{5}{3}-\frac{\sqrt{3}}{2\pi}}$	1.667	$\sqrt{3}\sqrt{1+\frac{\sqrt{3}}{2\pi}}$	1.956
Line current, rms value	no load	$\sqrt{2}$	1.414	$\sqrt{6}$	2.449
	maximum load	$\sqrt{2}\sqrt{\frac{5}{3}-\frac{\sqrt{3}}{2\pi}}$	1.667	$\sqrt{7+\frac{3\sqrt{3}}{2\pi}}$	2.798
Rating factor of one reactor		$\frac{(1+\sqrt{2})\sqrt{\frac{5}{3}-\frac{\sqrt{3}}{2\pi}}}{2\sqrt{6}\left(1-\frac{3\sqrt{3}}{4\pi}\right)}$	0.990	$\frac{(1+\sqrt{3})\sqrt{1+\frac{\sqrt{3}}{2\pi}}}{2\sqrt{3}\left(1+\frac{3\sqrt{3}}{2\pi}\right)}$	0.420
Total rating factor of circuit			2.970		2.572
Effective power factor of circuit			0.287		0.464

All values are per unit for unity turns ratio on the reactors. Bases of per-unit quantities:
Voltage = E_m, maximum value of line-to-neutral voltage, volts
Current = I_d, direct current bias, amperes
Resistance = E_m/I_d
Power = E_mI_d

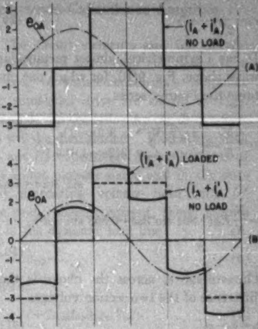


Fig. 11. Wave forms of line current in symmetrical frequency tripler

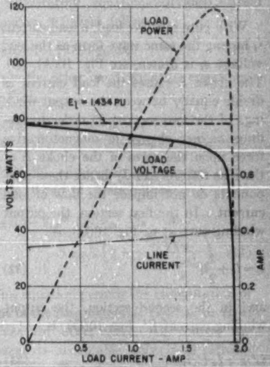


Fig. 12. Performance characteristics of symmetrical frequency tripler

frequency tripler with filtered output is three times the value of equation 27

$$\sum F=3F=3.372 \tag{28}$$

By way of comparison, the total rating factor of a similar type of frequency doubler with filtered output is approximately 2; see reference 2, equation 12.

PERFORMANCE DATA

Performance characteristics of an experimental basic frequency tripler are shown in Fig. 5, which has been obtained from reference 6, and may be considered typical for this circuit. The saturable reactor specifications are as follows:

Core: Arnold Engineering Company toroidal tape-wound core, Catalogue no. 5468-D2, 0.002-inch Deltamax tape (grain-oriented 50 per cent nickel, 50 per cent iron alloy), inside diameter 2.5 inches, outside diameter 3.5 inches, height 1.0 inch.

Windings: Input winding 2,550 turns 0.0201-inch wire, output winding 380 turns 0.0508-inch wire, bias winding 152 turns 0.0359-inch wire.

Supply: 450 volts, 3-phase, 60 cps, bias current 2.5 amperes

Note that a separate bias winding was used. In the experimental circuit, the choke was considerably larger than would normally be used.

Symmetrical Magnetic Frequency Tripler

This circuit, Fig. 6, has two sections, each consisting of a group of saturable reactors of the same type used in the basic frequency tripler. The output windings of the two sections are connected in series opposition, and carry a d-c I_d which biases the two groups of reactors with opposite polarity. The method of biasing shown in Fig. 6, consisting of the bias supply transformer T and bias rectifier *REC*, has been found convenient in practice. No direct current will flow through the load R if the sections are perfectly balanced. The choke L has two equal windings which are closely coupled.

STEADY-STATE OPERATION WITH NO LOAD

The first section of the circuit, in which the reactors are biased with conventionally positive polarity, develops the wave forms of Fig. 7 under no-load conditions. These wave forms are identical with those of Fig. 3 for the basic frequency tripler with positive bias. The reactors in the second section of the circuit are biased with conventionally negative polarity, and they generate wave forms as shown in Fig. 8, which are phase-shifted by π radians and reversed in polarity with respect to the wave forms produced by the first section.

The voltages e_o and e_o' developed in the output windings of the two sections are shown in Fig. 9(A). By the action of the choke L, the voltage e_l appearing across the load R is the instantaneous average of the voltages e_o and e_o'

$$e_l=\frac{1}{2}(e_o+e_o') \tag{29}$$

It has a symmetrical, almost rectangular wave shape, Fig. 9(B), for which one obtains the Fourier series

$$e_l=\frac{N_o}{N_i}E_m\frac{9\sqrt{3}}{\pi}\sum_{n=1}^{\infty}\frac{3(2n-1)}{[3(2n-1)]^2-1}\times \sin 3(2n-1)\omega t$$

$$=\frac{N_o}{N_i}E_m(1.861 \sin 3\omega t+0.558 \sin 9\omega t+\ldots) \tag{30}$$

The voltage e_c across the choke is the difference of the two section voltages

$$e_c=e_o-e_o' \tag{31}$$

This voltage e_c contains no harmonics of lower order than the sixth, and has a saw-toothed wave shape; see Fig. 9(C).

LOAD CONDITION

With pure resistive load, a load current i_l having the same wave form as the load voltage e_l is drawn, see Fig. 10(A), (B). The choke L causes the load current to divide equally between the output windings of the two sections in order to produce no net change in magnetomotive force upon the cores of the choke L or the bias transformer T; hence these components do not impede the flow of load current. In the first section, the output winding current i_o, Fig. 10(C), is

$$i_o=I_d+\frac{1}{2}i_l \tag{32}$$

and in the second section, the output winding current i_o', Fig. 10(G), is

$$i_o'=-I_d+\frac{1}{2}i_l \tag{33}$$

These currents are reflected into the input windings of the six reactors as illustrated in Fig. 10(D), (E), (F), and Fig. 10(H), (I), (J), respectively. The total current drawn from phase A of the supply, Fig. 11, is the sum of the currents drawn by phase A of each section; the currents in phases B and C are similar. It is seen that the line currents possess half-wave symmetry.

Maximum power output into the load is obtained when the peak value of the load current equals twice the direct bias current, corresponding to a critical value of load resistance. When the load resistance is smaller than the critical value, a change in the mode of operation occurs and the load voltage drops abruptly; see Fig. 12. The circuit is effectively a source of constant triple-frequency current with rectangular wave shape of magnitude $2\times I_d$ when the load impedance is small.

PERFORMANCE DATA

The rating of the symmetrical frequency tripler may be calculated by the same method as already applied to the basic frequency tripler. Data for the case of unfiltered output into resistive load are given in Table I. Performance characteristics of a typical circuit[5] are shown in Fig. 12. The saturable reactors are the same as were used in the basic frequency tripler experimental unit. It is observed that the symmetrical tripler is superior to the basic tripler in performance, output wave form and rating factor. However, the basic tripler is economical in the smaller sizes.

Discussion

With sharply saturating core material the actual wave shapes produced in a frequency multiplier circuit are very close to the ideal derived by means of the simplified analysis. For design purposes the actual voltages and currents in a circuit may be predicted with normally sufficient accuracy from the ideal by application of empirical "loss" factors, depending upon the core material, if its characteristic does not differ greatly from Fig. 1.

Magnetic frequency multipliers tend to have relatively high regulation, due to leakage inductance and the after-saturation inductance of the nonideal reactors, which may be minimized by careful design. A series capacitor may be used to resonate with the effective inductance of the unit, improving regulation. In some applications, a tuned filter may be used to eliminate undesirable harmonics. This modifies the currents in the circuit under loaded conditions and changes the rating factor somewhat, as shown in the case of the basic tripler.

There are various possible modifications of the circuit arrangements described, which do not change the fundamental mode of operation. For example, separate output and bias windings may be provided, as in the case of the experimental units, to isolate the d-c bias supply from the load. The rating factors of such arrangements will, in general, differ from the values derived here. In addition, a number of constructional variations are possible, such as making the output winding common to several cores, like an amplistat control winding.

References

1. Mercury Arc Power Rectifiers (book), O. K. Marti, H. Winograd. McGraw-Hill Book Company, Inc., New York, N. Y., 1930, pp. 176–86.

2. Magnetic Frequency Conversion, L. C. Harriott. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 9, 1953, pp. 78–87.

3. Magnetic Frequency Multipliers, L. J. Johnson, S. E. Rauch. *AIEE Transactions*, vol. 73, pt. I, Nov. 1954, pp. 448–52.

4. Analysis and Design of a Magnetic Frequency Multiplier, Otto J. M. Smith, Jalal T. Salihi. *Ibid.*, vol. 74, pt. I, Mar. 1955, pp. 99–106.

5. Differential Analyzer Study of Harmonic Power Generation with Nonlinear Impedance Element, Paul E. Russell, H. A. Peterson. *Ibid.*, vol. 70, pt. I, 1951, pp. 917–20.

6. An Introduction to the Theory of Magnetic Frequency Multipliers Using Biased Magnetic Cores, D. W. Leiby. *M.S. Thesis.* Union College, Schenectady, N. Y., 1955.

Optimum Design of Common-Emitter Transistor Audio Amplifiers

L. M. VALLESE
ASSOCIATE MEMBER AIEE

Synopsis: An analytical procedure of design of common-emitter transistor audio amplifiers is given. The conditions of optimum design for maximum power output are derived for resistance-capacitance and transformer-coupled amplifiers, taking in consideration the limitations of the collector characteristics. Finally, the distortion caused by nonlinearity of the collector characteristic is computed.

THE problem of the design of amplifiers for maximum power output is of great importance for transistors. Differently from vacuum-tube amplifiers which have theoretically infinite power gain but finite power output, transistor amplifiers have finite power gain and finite power output. For this reason it is often important to maximize them for power output at least in the last two stages. Such design is obtained with consideration of the maximum limitations of the collector electric characteristics; in particular it is required to determine the quiescent point of operation so that maximum possible input drive may be obtained with the available linear range and the required stabilization of the operating point. At present the analysis of these conditions is performed graphically, following a cut-and-trial procedure which requires specialized skill.[1,2]

In the following an analytical procedure of design, based on a linear approximation of the characteristics, will be given. Its aim is to provide a systematic approach and to emphasize the interdependence of the variables involved and the degrees of freedom available to the designer. A similar procedure could be developed for pentode amplifiers, the design of which is generally obtained graphically. It will be assumed that the short-circuit current gain α_{cb} is independent of the emitter current. Only the common-emitter configuration will be considered in detail in this paper, since the same procedure applies to the other configurations. The problem of temperature stabilization will not be considered and it will be assumed that the transistor characteristics are given in condition of temperature steady state. In practice it will be necessary to apportion a suitable part of the linear region for thermal variations.

In Fig. 1 a typical example of input and output characteristics of common-emitter low-power p-n-p transistor is shown. For linear amplification, these characteristics should consist of parallel straight lines; in the following it will be assumed that within certain ranges such is the case. In a later section the problem of the distortion will be discussed.

The linearized input and output characteristics may be represented with the equations

$$I_b=g_{bb}(V_{be}-V_{be}^0)+g_{bc}V_{ce} \tag{1}$$

if

$$|I_b|\geq|I_{b\,\min}|$$

$$|V_{be}\geq|V_{be\,\min}|$$

and

$$I_c=I_0'+\alpha_{cb}(I_b-I_b')+g_{ce}V_{ce} \tag{2}$$

if $|I_b|\geq|I_b'|\geq|{}_bI_{b\,\min}|$. The boundaries of the linear region in the collector plane are

$$|V_{ce}'|\leq|V_{ce}|\leq|V_{ce\,\max}| \tag{3}$$

$$|I_c'|\leq|I_c|\leq|I_{c\,\max}| \tag{4}$$

$$V_{ceQ}I_{cQ}\leq P_c \tag{5}$$

The explanation of the various symbols is given in Appendix I. It should be noted that, in most cases, the parameter g_{bc} is practically negligible in a considerable range of the variables. The design will be carried out only in the collector plane.

Resistance-Capacitance Amplifier

The circuit elements and the electric quantities of interest in the design of a typical resistance-capacitance amplifier for sinusoidal drive are shown in Figs. 2 and 3. In the following the drop across G_3 will be disregarded. Indicating with I_{bs} and E_0 respectively the amplitudes of the input current signal and of the output alternating voltage, and with I_c'' the peak collector current, there follows from the dynamic load line

$$I_c''=2\,I_{cQ}-I_c' \tag{6}$$

where

$$I_c''=I_0'+2\alpha_{cb}I_{bs}+g_{ce}V_{ce}' \tag{7}$$

$$I_{cQ}=(G_2+G_L)E_0+I_c' \tag{8}$$

$$I_{cQ}=G_2(E_{cc}-V_{ce}'-E_0) \tag{9}$$

$$I_c'=I_0'+g_{cc}(V_{ce}'+2E_0)$$

Substituting equations 7 and 8 in equation 6, there follows

$$\alpha_{cb}I_{bs}=(G_2+G_L+g_{ce})E_0 \tag{10}$$

On the other hand, equating equations 8 and 9

$$G_2(E_{cc}-V_{ce}')-g_{ce}V_{ce}'-I_0'=(2G_2+2g_{ce}+G_L)E_0 \tag{11}$$

Finally, eliminating E_0 between equations 10 and 11, there follows

$$G_2(E_{cc}-V_{ce}')-g_{ce}V_{ce}'-I_0'=\alpha_{cb}I_{bs}\left(1+\frac{G_2+g_{ce}}{G_2+g_{ce}+G_L}\right) \tag{12}$$

This equation permits the computation of one of the parameters G_2, G_L, E_{cc}, or I_{bs} if given the other three. Of course, the design is acceptable only if it satisfies the limitations of equations 3, 4, and 5.

The design for maximum power output generally is of academic interest because of the large d-c supply voltage required and of the low collector efficiency involved; however, it is important as a reference optimum condition. Two separate cases must be considered, i.e., the design for given I_{bs} and that for given E_{cc}; it should be noticed that the first case also corresponds to maximum power gain, since the input power is $I_{bs}^2/2G_{in}$.

When I_{bs} is given, expressing E_0 with equation 10 and maximizing the quantity $P_L = G_LE_0^2/2$, G_2 is found as small as possible and

$$G_L=G_2+g_{ce} \tag{13}$$

Correspondingly, the maximum power output and the related power gain are

$$P_{L\,\max}=\alpha_{cb}^2I_{bs}^2/8G_L$$

$$A_P=\alpha_{cb}^2G_{in}/4G_L \tag{14}$$

Paper 55-667, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the National Electronics Conference, Chicago, Ill., October 3–5, 1955, and re-presented at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted March 28, 1955; made available for printing December 2, 1955.

L. M. VALLESE is with Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

The work described has been done at the Microwave Research Institute of the Polytechnic Institute of Brooklyn, under a contract sponsored by the Office of Naval Research.

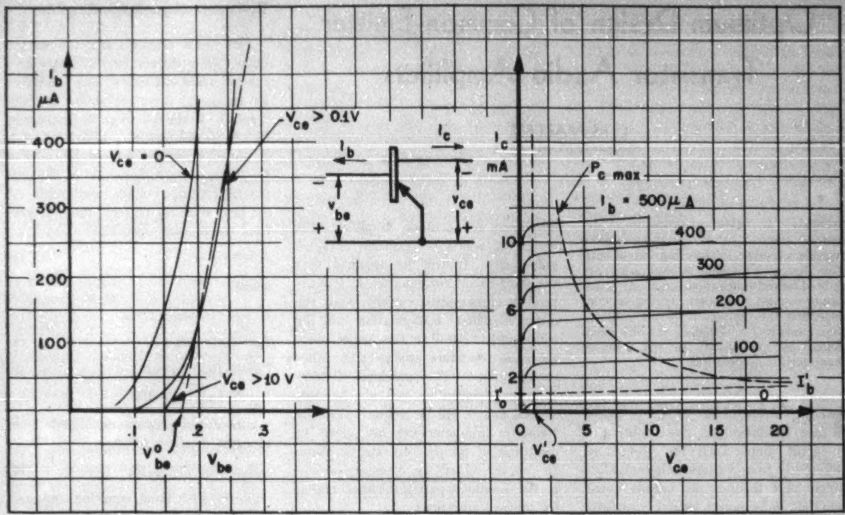


Fig. 1. Example of input and output characteristics of low-power common-emitter transistor

The lowest value of G_2 is obtained considering that the limit position of the quiescent point is at the intersection of the I_c characteristic corresponding to $I_b = I_{bQ} = I_b' + I_{bs}$ with either the P_c hyperbola or with the vertical line

$$V_{ce} = V_{ceQ} = (V_{ce\ max} + V_{ce}')/2$$

whichever is smaller.

In the first case from

$$I_{cQ} = I_0' + \alpha_{cb} I_{bs} + g_{cc} V_{ceQ} = P_c / V_{ceQ}$$

there follows

$$V_{ceQ} = -\frac{I_0' + \alpha_{cb} I_{bs}}{2 g_{cc}} + \sqrt{\frac{(I_0' + \alpha_{cb} I_{bs})^2}{4 g_{cc}^2} + \frac{P_c}{g_{cc}}} \quad (15)$$

In particular, if g_{cc} is neglected, it is found approximately in excess

$$V_{ceQ} \cong \frac{P_c}{I_0' + \alpha_{cb} I_{bs}} \quad (16)$$

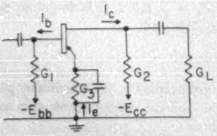


Fig. 2. Common-emitter resistance-capacitance amplifier

From the value of V_{ceQ}, $E_0 = V_{ceQ} - V_{ce}'$ is obtained, and from equation 10 is obtained

$$G_L = \frac{\alpha_{cb} I_{bs}}{2 E_0}$$

$$G_2 = G_L - g_{cc} \quad (17)$$

Upper values of P_L are $P_c/4$ or $(\alpha_{cb} I_{bs} V_{cemax})/8$ according if $I_{bs} \gtrless [(2P_c)/(\alpha_{cb} V_{cemxa})]$

Finally, E_{cc} follows from equation 12

$$E_{cc} = V_{ce}' + \frac{1}{G_2}\left(\frac{3}{2}\alpha_{cb} I_{bs} + g_{cc} V_{ce}' + I_0'\right)$$

Consider now the optimum design for given E_{cc}. Expressing E_0 from equation 11 and maximizing P_L, it is found that G_2 should be as large as possible and

$$G_L = 2(G_2 + g_{cc}) \quad (18)$$

As a first approximate uppermost value for G_2, a value of $4P_c/E_{cc}^2$ for which the static load line is tangent to the P_c hyperbola is assumed. Correspondingly, I_{bs} follows from equation 12, which with the substitution into equation 18 reads

$$I_{bs} = \frac{3}{4\alpha_{cb}} (G_2(E_{cc} - V_{ce}') - g_{cc} V_{ce}' - I_0') \quad (19)$$

These values of G_2 and I_{bs} are acceptable if

$$I_{bs} \leq (I_{c\ max} - I_0' - g_{cc} V_{ce}')/2\alpha_{cb} \quad (20)$$

If the latter condition is not satisfied, I_{bs} will be taken equal to the right-hand member of equation 20 and G_2 will follow from equation 19.

In first approximation, the design will be limited by the $I_{c\ max}$ value or by the P_c value, Fig. 4, depending on whether

$$E_{cc} > 6P_c/I_{c\ max} \text{ or } E_{cc} < \frac{6P_c}{I_{c\ max}}$$

From the values of I_{bs} and G_L there follows that $E_0 = 2\alpha_{cb} I_{bs}/3G_L$ and, correspondingly, the maximum power output and the related power gain are

$$P_{L\ max} = \frac{2\alpha_{cb}^2 I_{bs}^2}{9 G_L}$$

$$A_P = \frac{4}{9} \alpha_{cb}^2 \frac{G_{in}}{G_L}$$

Approximate upper values of $P_{L\ max}$ are $P_c/4$ for the P_c limited case and $E_{cc} I_{c\ max}/24$ for the $I_{c\ max}$ limited case.

The design of the input circuit follows simply after the design of the output circuit has been obtained. The value of E_{cc} must be adjusted to take into account the drop of G_3.

Transformer-Coupled Amplifier

The procedure of design of the transformer coupled amplifier, Fig. 5, is similar to that developed for the resistance-

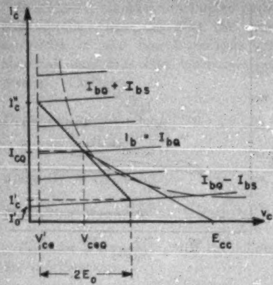


Fig. 3. Static and dynamic load lines of common-emitter resistance-capacitance amplifier

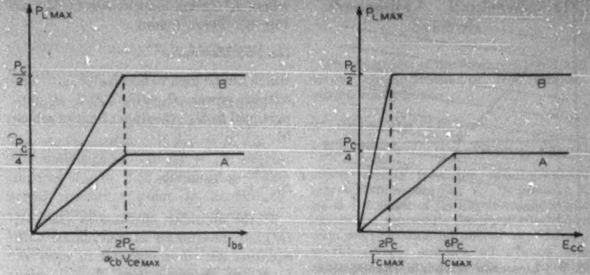


Fig. 4. Upper boundaries of $P_{L\,max}$ for given I_{bs} and for given E_{cc}

A—Resistance-capacitance amplifier
B—Transformer-coupled amplifier

capacitance amplifier, but results in simpler equations since

$$V_{ceQ} = E_{cc} = E_0 + V_{ce}'$$

Indicating with G_L', the primary transformed load admittance, there results from the dynamic load line in the collector plane

$$I_c'' = 2I_{cQ} - I_c'$$

where

$$I_c'' = I_0' + 2\alpha_{cb}I_{bs} + g_{cc}V_{ce}'$$

$$I_c' = I_0' + g_{cc}(V_{ce}' + 2E_0)$$

$$I_{cQ} = G_L'E_0 + I_0' + g_{cc}(V_{ce}' + 2E_0)$$

Substituting, it is found that

$$\alpha_{cb}I_{bs} = (G_L' + g_{cc})E_0 = (G_L' + g_{cc})(E_{cc} - V_{ce}') \tag{21}$$

Equation 21 permits the computation of one of the parameters G_L', E_{cc}, or I_{bs} if the other two are given. The resulting design is acceptable provided it satisfies the limitations of equations 3, 4, and 5.

The design for maximum power output may be considered again in two separate cases, i.e., for given I_{bs} and for given E_{cc}; in the first case, maximum power gain is also obtained.

For given I_{bs}

$$P_L = \frac{G_L'}{2}\,\frac{\alpha_{cb}^2 I_{bs}^2}{(G_L' + g_{cc})^2}$$

and maximum power output is obtained when $G_L' = g_{cc}$. However, this condition is not practical because, in general, it is violative of the limitations of equations 3 and 5 and because g_{cc} is too small. On the other hand, from equation 21 is obtained

$$G_L' = \frac{\alpha_{cb}I_{bs}}{E_{cc} - V_{ce}'} - g_{cc} \tag{22}$$

so that the lowest value of G_L' may be determined as a function of $E_{cc} = V_{ceQ}$. The latter is either the value of equation 15 or $(V_{ce\ max} + V_{ce}')/2$, whichever is smaller. In first approximation, E_{cc} is limited by the P_c value or by the $V_{ce\ max}$ value if

$$\alpha_{cb}I_{bs} \gtrless 2P_c/V_{ce\ max}$$

The power output and the related power gain are

$$P_{L\max} = \frac{1}{2}\left(\frac{\alpha_{cb}I_{bs}}{E_{cc} - V_{ce}'} - g_{cc}\right)(E_{cc} - V_{ce}')^2$$

$$A_P = \frac{2G_{in}}{I_{bs}^2}\,P_{L\ max}$$

Approximate upper values of $P_{L\max}$ are $P_c/2$, when E_{cc} is computed with equation 15, and $\alpha_{cb}I_{bs}V_{ce\ max}/4$ in the alternate case.

For given E_{cc}, the power output is

$$P_L = G_L'(E_{cc} - V_{ce}')^2/2$$

and maximum power output is obtained for G_L' as large as possible. The uppermost value of G_L' is obtained from equation 22, where I_{bs} is the smaller of

$$\frac{1}{\alpha_{cb}}\left(\frac{P_c}{E_{cc}} - I_0' - g_{cc}(E_{cc} - V_{ce}')\right)$$

or

$$\frac{1}{2\alpha_{cb}}\,(I_{c\ max} - I_0' - g_{cc}V_{ce}') \tag{23}$$

Approximate upper values of $P_{L\max}$, Fig. 4, are respectively for the first and the second case of equation 23, i.e., for $E_{cc} \gtrless [(2P_c)/(I_{c\ max})]$

$$P_c/2$$

$$E_{cc}I_{c\ max}/4$$

In completing the design, the value of E_{cc} should be adjusted to take into account the drop across G_3.

Transformer-Coupled Push-Pull Amplifier

Likewise, in the case of vacuum tubes,[3] the procedure of design of transformer-coupled push-pull amplifiers is simplified with the introduction of the composite transistor. The latter is defined as the linear equivalent transistor which has the same gain and power output as the push-pull amplifier, Fig. 6, but has an

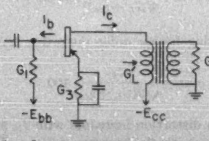


Fig. 5. Common-emitter transformer-coupled amplifier

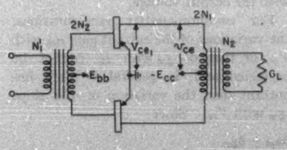


Fig. 6. Common-emitter push-pull amplifier

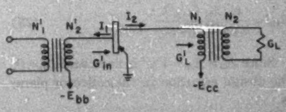


Fig. 7. Composite transistor amplifier

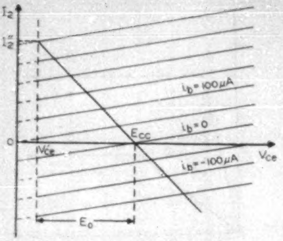


Fig. 8. Linearized composite characteristics of push-pull common-emitter amplifier

input transformer of ratio N_1' to N_2' and an output transformer of ratio N_1 to N_2, with the same source and load admittances; see Fig. 7. If the currents and voltages of the two transistors of the push-pull amplifier are indicated respectively with the subscripts 1 and 2, the corresponding currents and voltages of the composite transistor are

$$I_1 = I_{b1} - I_{b2},\ V_1 = V_{be1}$$

$$I_2 = I_{c1} - I_{c2},\ V_2 = V_{ce1}$$

Indicating with i_b, v_{be}, i_c, and v_{ce} the instantaneous values of the alternating currents and voltages

$$I_{b1} = I_{bQ} + i_b,\ I_{b2} = I_{bQ} - i_b$$

$$V_{be1} = E_{bb} + v_{be},\ V_{be2} = E_{bb} - v_{be} = 2E_{bb} - V_{be1}$$

$$V_{ce1} = E_{cc} - v_{ce},\ V_{ce2} = E_{cc} + v_{ce} = 2E_{cc} - V_{ce1}$$

The linearized composite input and output characteristics are

$$I_1 = k_{11}(V_{be1} - E_{bb}) + k_{12}v_{ce} \tag{24}$$

$$I_2 = k_{21}i_b + k_{22}(V_{ce1} - E_{cc})$$

where, for class-A operation, $k_{11} = 2g_{bb}$, $k_{12} = 2g_{bc}$, $k_{21} = 2\alpha_{cb}$, and $k_{22} = 2g_{ce}$, and for class-B operation $(AB-B)$, $k_{11} \cong g_{bb}$, $k_{12} \cong g_{bc}$, $k_{21} \cong \alpha_{cb}$, and $k_{22} \cong g_{cc}$. Equations 24 ordinarily are found by graphical construction after the quiescent point of operation of the component transistors has been determined, as shown in the previous section.

The slope of the composite load line in the collector plane is $-G_L' = -G_L(N_2/N_1)^2$, while that of the input dynamic characteristic in the base plane is

$$G_{in}' = \frac{k_{11}(k_{22} + G_L')}{k_{12}k_{21} + k_{22} + G_L'} \tag{25}$$

Since $E_0 = E_{cc} - V_{ce}'$ is fixed and $I_{2Q} = 0$, the design equation in the collector plane is

$$I_2'' = G_L'E_0 = k_{21}I_{bs} + k_{22}(V_{ce}' - E_{cc})$$

where I_2'' is the peak value of I_2; see Fig. 8. There follows

$$(k_{22} + G_L')E_0 = k_{21}I_{bs}$$

which can be used to find G_L' or I_{bs}. The output power $P_L = (G_L'E_0^2)/2$ is proportional to G_L', the upper limit of which is

$$\frac{k_{21}}{E_{cc} - V_{ce}'} I_{bs\,\max} - k_{22}$$

If the amplifier operates in class A, $I_{bs\,\max}$ is determined in accordance with equation 23; if it operates in class B, $I_{bs\,\max}$ is obtained from the following equation

$$I_{bs\,\max} = \frac{1}{\alpha_{cb}}(I_c - I_0' - g_{ce}V_{ce}')$$

where I_c is the smaller of $I_{c\,\max}$ and $P_c/(1/\pi - 1/4)E_{cc}$. Correspondingly, the approximate upper limits of the output power are, in class A, for $E_{cc} \lessgtr 2P_c/I_{c\,\max}$

$$\frac{1}{2} I_{c\,\max}E_{cc} \text{ or } P_c$$

and, in class B, for $E_{cc} \lessgtr P_c/(1/\pi - 1/4) I_{c\,\max}$

$$\frac{1}{2} I_{c\,\max}E_{cc} \qquad P_c/\left(\frac{2}{\pi} - \frac{1}{2}\right) = 7.3P_c$$

Finally, the input admittance of the primary input transformer is $(N_1'/N_2')^2 G_{in}'$, where G_{in}' is given by equation 25.

Collector Distortion

Distortion occurs in power transistor amplifiers because of nonlinearities of the input and the output characteristics. Although these are not additive, they have been investigated separately[1] in order to arrive at simple approximate expressions. The base distortion has been obtained with the assumption that the input admittance varies directly with I_e; the collector distortion (excluding the presence of undesirable clipping) has been obtained expressing I_c as a power series and assuming that the current gain is invariant with changes in operating voltage and current, and the slope of the output characteristics varies with the output voltage.

For common-emitter configurations, the variations of the current gain α_{cb} with V_{ce}, or that of g_{ce} with I_b, cannot be neglected and, in general, is more important than the variation of the slope g_{ce} with V_{ce}. Since

$$\frac{\partial \alpha_{cb}}{\partial V_{ce}} = \frac{\partial g_{ce}}{\partial I_b}$$

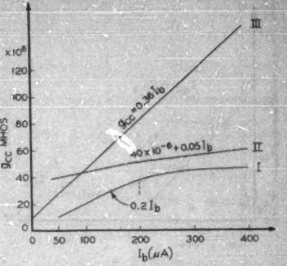


Fig. 9. Plot of g_{ce} versus I_b for three different transistors

the latter quantity may be obtained from a plot of g_{ce} versus I_b; see Fig. 9.

It is important to evaluate the collector distortion and to take into account the previous considerations. The power series expansion of I_c about the quiescent point may be written with the use of small letters for incremental variations, letting $\gamma = (\partial\alpha_{cb})/(\partial V_{ce})$ and neglecting $(\partial g_{ce})/(\partial V_{ce})$

$$i_c \cong \alpha_{cb}i_b + g_{ce}v_{ce} + \gamma v_{ce}i_b \tag{26}$$

where α_{cb}, g_{ce}, and γ are evaluated at the quiescent point.

In the case of a single transformer-coupled amplifier

$$i_c = -G_L'v_{ce}$$

and therefore

$$v_{ce} = \frac{-\alpha_{cb}i_b}{G_L' + g_{ce} + \gamma i_b}$$

Expanding in power series and assuming that $i_b = I_{bs}\cos\omega t$, there follows

$$v_{ce} = -\frac{\alpha_{cb}i_b}{G_L' + g_{ce}}\left(1 - \frac{\gamma i_b}{G_L' + g_{ce}} + \frac{\gamma^2 i_b^2}{(G_L' + g_{ce})^2} - \dots\right) =$$

$$-\frac{\alpha_{cb}i_b}{G_L' + g_{ce}}\left[\left(1 + \frac{3\gamma^2 I_{bs}^2)}{4(G_L' + g_{ce})^2}\right) \times \cos\omega t - \frac{\gamma I_{bs}}{2(G_L' + g_{ce})} - \frac{\gamma I_{bs}}{2(G_L' + g_{ce})} \times \cos 2\omega t + \frac{\gamma^2 I_{bs}^2}{4(G_L' + g_{ce})^2}\cos 3\omega t + \dots\right]$$

Neglecting higher order distortion terms

$$D_{0,2} \text{ per cent} \cong \frac{\gamma I_{bs}}{\sqrt{2}(G_L' + g_{ce})} 100$$

i.e., the distortion increases with I_{bs} and decreases with the increase of G_L' and of g_{ce}.

In a class-*A* transformer-coupled push-pull amplifier, assuming that the characteristics of the two transistors are identical, only odd-order harmonics are present. Therefore

$$D_3 \text{ per cent} \cong \frac{\gamma^2 I_{bs}^2}{4\left(g_{cc}+\dfrac{G_L'}{2}\right)^2}\,100$$

where $G_L = (N_2/N_1)^2 G_L$.

Numerical Example

Consider a transistor with the following specifications

$g_{bb}=3\times10^{-3}$ mhos, $g_{cc}=50\times10^{-6}$ mhos, $\alpha_{cb}=20$

$g_{bc}=0$, $I_b'=20$ microamperes, $I_0'=0.5$ milliampere, $I_{c\,max}=20$ milliamperes

$V_{ce}'=1$ volt, $V_{ce\,max}=40$ volts, $P_c=100$ milliwatts (at steady-state temperature)

Design a resistance-capacitance and a transformer-coupled amplifier for maximum power output (and maximum power gain) with given I_{bs} = 200 microamperes. With the resistance-capacitance configuration it is found that

$V_{ceQ}=18.5$ volts (equation 15), $E_c=17.5$ volts, $I_{cQ}=5.4$ milliamperes

$G_L=114\times10^{-6}$ mhos ($R_L=8{,}700$ ohms)

$G_2=64\times10^{-6}$ mhos ($R_2=15{,}600$ ohms)

$E_{cc}=102.5$ volts, $P_L=17.5$ milliwatts

It is noted that, if g_{cc} is neglected, equation 16 provides V_{ceQ} = 22 volts, with an error of 20 per cent in excess. The limit value of V_{ceQ} according to equation 3 is 20.5 volts.

With the transformer-coupled configuration

$V_{ceQ}=18.5$ volts, $E_c=17.5$ volts, $I_{cQ}=5.4$ milliamperes

$G_L'=178\times10^{-6}$ mhos ($R_L'=5{,}600$ ohms)

$E_{cc}=18.5$ volts, $P_L=27.4$ milliwatts

Consider now a push-pull configuration in class-*B* (*AB B*), with E_{cc} = V_{ceQ} = 18.5 volts. If I_{bQ} = 0, I_{cQ} = 0.05 milliampere; if I_{bQ} = 20 microamperes, I_{cQ} = 0.55 milliampere; the actual operating point is determined experimentally since the characteristics of the two transistors are never identical.

If I_{bs} is taken equal 400 microamperes

$E_c=17.5$ volts, $G_L'=406\times10^{-6}$ mhos, $P_L=62$ milliwatts

On the other hand the upper limit of I_s is $I_{c\,max}=20$ milliamperes so that the upper limit of I_{bs} is 0.97 milliampere. Correspondingly

$G_L'=1.14\times10^{-3}$ mho, $P_L=175$ milliwatts

Note that, if the upper limit of I_s could be taken equal $P_c/(1/\pi-1/4)E_{cc}=79$ milliamperes, a value $P_L=730$ milliwatts would result. The over-all and the input distortion are not considered here, but the collector distortion defined in the previous section will be evaluated with the assumption that $g_{cc}=0.2\ I_b$. In this case, for a single transformer-coupled amplifier

$$D_{0,2} \text{ per cent} \cong \frac{0.2\times200}{\sqrt{2}(178+0.2\times200)}\,100 = 13 \text{ per cent}$$

while for a class-*A* push-pull amplifier with I_b = 200 microamperes

$$D_3 \text{ per cent} \cong \frac{1}{4}\left(\frac{0.2\times200}{178+0.2\times200}\right)^2 100 = 0.8 \text{ per cent}$$

Appendix I. Explanation of Symbols

The equations of the linearized input and output characteristics of the common-emitter configuration

$$I_b=g_{bb}(V_{be}-V_{be}^0)+g_{bc}V_{ce} \qquad (1)$$

$$I_c=I_0'+\alpha_{cb}(I_b-I_b')+g_{cc}V_{ce} \qquad (2)$$

follow directly from the usual graphical representation; see Fig. 1. In equation 1, g_{bb} is the slope of the input characteristic curves and represents the input admittance with short-circuited output circuit, while g_{bc} is the reverse transfer admittance with short-circuited input circuit. In equation 2, α_{cb} is the forward current gain with short-circuited output circuit and g_{cc} is the slope of the output characteristic curves, i.e., the output admittance for the open input circuit. In addition, I_b' is the minimum base current consistent with the desired linearity ($|I_b'|\geq|I_{b\,min}|$), I_0', and V_{ce}' are shown in Fig. 1, and $I_0'-\alpha_{cb}I_b'$ is the collector current for $I_b=0$, $V_{ce}=0$. Finally, $I_c'=I_0'+g_{cc}V_{ce}$ is the collector current for $I_b=I_b'$.

In terms of the hybrid parameters of the common emitter (for large signal)

$$g_{bb}=\frac{1}{h_{11}},\quad \alpha_{cb}=h_{21},\ g_{bc}=-\frac{h_{12}}{h_{11}},\ g_{cc}=h_{22}$$

and in terms of the open-circuit parameters

$$g_{bb}\cong\left(r_b+\frac{r_e}{1-\alpha_{ce}}\right)^{-1},\ g_{cc}=[r_c(1-\alpha_{ce})]^{-1}$$

$$g_{bc}=-\frac{r_e}{r_c(1-\alpha_{ce})}\left(r_b+\frac{r_e}{1-\alpha_{ce}}\right)^{-1}$$

References

1. Transistor Audio Amplifiers (book), R. F. Shea. John Wiley & Sons, Inc., New York, N. Y., 1955, chap. 7.

2. Transistor Circuits (book), R. F. Shea. *Ibid.*, 1953, chap. 7.

3. Applied Electronics (book), T. S. Gray. McGraw-Hill Book Company, Inc., New York, N. Y., 1954.

A Single-Transistor Magnetic-Coupled Oscillator

KAN CHEN
ASSOCIATE MEMBER AIEE

A. J. SCHIEWE
ASSOCIATE MEMBER AIEE

AN IMPORTANT oscillator circuit which converts d-c to a-c square waves having an output frequency proportional to the d-c input voltage was developed by Royer.[1] The circuit, shown in Fig. 1, employs a magnetic core constructed of square hysteresis loop material and two junction transistors operated as controlled switches.[2] As the circuit oscillates, the core flux swings repetitively between its positive and negative saturation levels, $\pm\phi_{max}$, the rate of flux excursion being proportional to the input voltage E_{in}. Thus, the frequency of oscillation is given by

$$f=\frac{E_{in}}{4N_1\phi_{max}} \qquad (1)$$

where N_1 is the number of turns of the windings in the collector circuit. Since the flux saturation levels ϕ_{max} vary with temperature, the output frequency is temperature-dependent. The output power is drawn from the input voltage source E_{in}, hence no power amplification is realized in this circuit.

Another oscillator circuit bearing a close resemblance to Royer's circuit was developed by Van Allen.[3] This circuit, shown in Fig. 2, employs two magnetic cores and two transistors, and produces also an a-c square wave having an output frequency proportional to the d-c input voltage. The general operating principle of Van Allen's circuit is similar to that of Royer's circuit.[3] However, there is a significant difference between the two circuits in consideration of the core flux excursion. Instead of swinging between saturation levels, the core flux in Van Allen's circuit is reset to an unsaturated value on alternate half-cycles in a manner similar to self-saturating magnetic amplifiers. This fact imparts to Van Allen's circuit the following advantages which justify the use of an additional core as compared to Royer's circuit:

1. The resetting action, and hence the frequency of oscillation, may be controlled by a current in an additional control winding at a low power level. Amplification of power is therefore achieved. The circuit offers the possibility of being operated as a frequency-modulated magnetic amplifier.

2. The additional means of controlling frequency provided by this current source facilitates frequency drift compensation.

3. The variation of flux saturation levels with temperature no longer has effect upon the frequency of oscillation, although variations of other core properties do.

The authors recognized the merits of Van Allen's circuit but were not completely satisfied with the explanation of the circuit operation given by Van Allen a critical study of which, substantiated by experimental results, precipitated the discovery that one half of Van Allen's or Royer's circuit would oscillate. The new oscillator circuit thus uses only one transistor and one core and retains the aforementioned merits of Van Allen's circuit. A fuller explanation of Van Allen's circuit and the description of the single-transistor magnetic-coupled oscillator follow.

Discussion of Van Allen's Circuit

Van Allen explains that the magnetic coupling of the two cores in his circuit permits one core to obtain a resetting voltage from the other core, which is being driven towards saturation. He is aware that the resetting core is not completely reset by the firing core because of the voltage drop across the rectifiers and winding resistances. He explains that these drops will determine to some extent the zero control current frequency. This is obviously true if one assumes that the circuit will sustain itself in oscillation. However, if these voltage drops were not compensated for, the cores would be reset less and less each cycle; hence, the circuit would eventually cease to oscillate for the conditions of zero or positive control current.

Experimental results obtained by the authors with the circuit of Fig. 2 lead to an explanation of the reset problem, an explanation based upon the factual existence of stray winding and transistor capacitances. It is believed that these stray capacitances, in particular those associated with the core windings, play a vital role in the mechanics of the resetting operation. Consideration of these capacitances leads to the following explanation of the operation of Van Allen's circuit for the condition of zero control current, and the cases for positive or negative control current follow directly from it.

Assume one of the transistors, Fig. 2, is conducting. The corresponding core is driven towards saturation, and at saturation a large surge of collector current takes place and, after this occurs, the transistor cuts off. This abrupt decrease of currents in the windings results in a large induced voltage (negative with respect to the induced voltage during the saturation period) across the core windings. If there were no stray capacitances in the circuit, the induced voltage would disappear as soon as the currents in the windings became zero. The resetting of the core would then depend entirely on the firing voltage available from the other core. However, the existing stray capacitances across the core windings and across the corresponding transistor tend to maintain this high induced voltage and cause the winding current to reverse direction. This action provides a negative exciting current and a proper volt-time area to reset the core partially. The frequency of oscillation will adjust itself so that the additional resetting action due to the stray capacitances is just enough to compensate for the loss due to voltage drops of part of the resetting volt-time area obtained from the firing core. The oscillograms in Fig. 3 show the large induced voltage across the windings at the beginning of the reset half-cycle. This induced voltage does not appear on the load because of the decoupling effect of the rectifiers.

The Single-Transistor Oscillator Circuit

With this reasoning, there appeared to be no reason why one half of Van Allen's circuit would not oscillate; see Fig. 4. That is, the core should be able to obtain sufficient resetting action for sustained oscillation from the stray capacitances, even for zero control current I_c. Of course, the frequency of oscillation would be determined by the magnitude of the stray capacitances and the magnitude of the load resistance. When this circuit was tested, the foregoing statements were

---

Paper 56-207, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956. Manuscript submitted June 6, 1955; made available for printing March 29, 1956.

Kan Chen is with Westinghouse Electric Corporation, East Pittsburgh, Pa., and A. J. Schiewe is with Purdue University, West Lafayette, Ind.

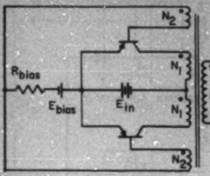


Fig. 1. Royer's circuit

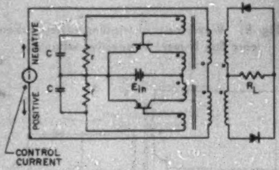


Fig. 2. Van Allen's circuit

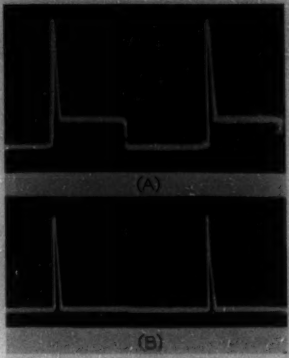


Fig. 3. Oscillograms pertaining to Van Allen's circuit (zero lines not shown in oscillograms)

A—Winding voltage
B—Rectifier voltage

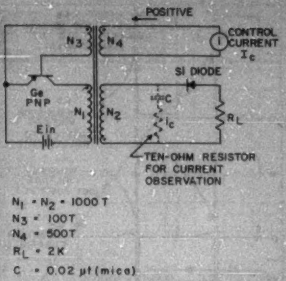


Fig. 4. Single-transistor magnetic-coupled oscillator with use of germanium p-n-p transistor

found to be true. The circuit oscillates without the rectifier in the load circuit. It is placed there to block the large negative induced voltage of the winding so that the output voltage will be rectangular. Winding voltage and rectifier voltage wave forms are shown in Figs. 5(A), 5(B), and 5(C). The negative spike on the load-voltage wave form appears because the Zener voltage of the rectifier has been exceeded. Since the firing voltage and the resetting voltage of the core are different in magnitude (and yet their volt-time areas must be equal), the firing and resetting periods are not equal.

The load voltage wave form may be made symmetrical by adjusting an external capacitance placed across any of the windings. The capacitance shown by dashed lines in Fig. 4 illustrates one possibility. The use of a large external capacitance has the accompanying advantages of swamping out the stray capacitances, which would be difficult to reproduce or to hold constant, and of reducing the peak negative induced voltage. Of course, the frequency of oscillation is lower than that of the stray capacitance case, because the amount of reset of the core flux is increased. The corresponding wave forms are shown in Figs. 6 and 7.

The theory of oscillation of this single-transistor circuit for zero control current may be more clearly stated in terms of the wave forms shown in Figs. 6 and 7. The load winding and the capacitor, Fig. 4, constitute a nonlinear tank circuit. The transistor is easily started into conduction by its own leakage through winding N_1 and the resulting positive voltage feedback action through the base winding N_3, if the load on the circuit is not excessive. Hence, the explanation of operation may be begun by assuming that the transistor is fully conducting. As the core flux reaches positive saturation and the transistor cuts off, the energy stored in the saturated inductance of the core is rapidly released, part being transferred to the capacitor and part being dissipated through core loss and resistance loss. This is revealed by the surge of capacitor current and the rapid reversal of the capacitor voltage, which reaches the negative peak in a very short time Δt, since the core remains saturated and the time constant is small during this period. As the capacitor forces energy back into the core, the core flux resets in the unsaturated level which results in a comparatively long time constant for this part of the transient. The damping of the tank circuit must be sufficiently small so that the winding voltage will overshoot and become positive again, Figs. 6(C) and 7(B), starting the next cycle of circuit oscillation by providing enough base drive for the transistor.

Of course, if the input voltage and the external capacitor are large enough, the energy stored in the capacitor at the beginning of the resetting half-cycle may be sufficient to reset the core to its negative saturation level. The winding voltage will then have a wave form as shown in Fig. 8(A). Once the core is saturated in the negative direction, the transistor is rendered conductive again by the fall-

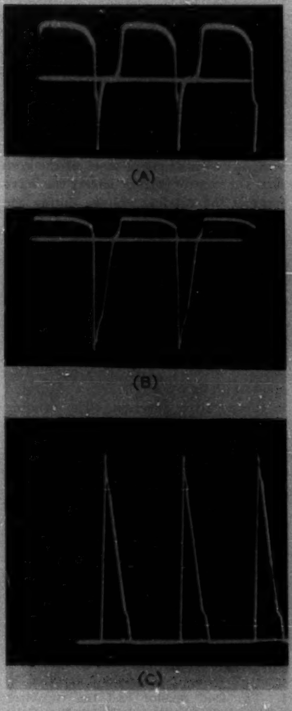


Fig. 5. Oscillograms pertaining to single-transistor magnetic-coupled oscillator without external capacitors (zero lines shown in oscillograms)

A—Load voltage
B—Winding voltage
C—Rectifier voltage

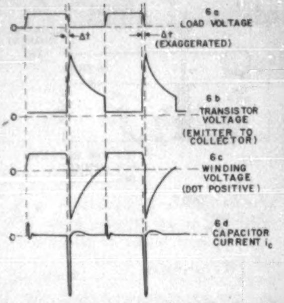


Fig. 6. Sketch of wave forms pertaining to single-transistor magnetic-coupled oscillator with external capacitor

back of the core flux to its negative remanence value, starting the next cycle of circuit oscillation.

When a control current I_c is present, the mechanics of core flux resetting operation is more complicated. In addition to varying the amount of reset of the core flux, the control current affects the wave form of the winding voltage during the resetting half-cycle through current transformer action. The negative winding voltage, which is also the capacitor voltage, will decay more or less rapidly, depending on whether the control current is positive (saturating) or negative (resetting). When a large negative control current is present, the magnitude of the negative winding voltage may actually increase with time until the core is saturated in the negative direction. The winding voltage will then have a wave form as shown in Fig. 8(B). It is therefore understandable that for some combination of values of the input voltage and the control current, the winding voltage may decay to zero before the core flux is reset to the negative saturation level. When this is the case, the transistor will remain nonconductive and oscillation of the circuit will cease.

In summary, the following two conditions have to be met in order that the single-transistor oscillator will commence and sustain oscillation:

1. The load resistor and the capacitor must not overload the oscillator during the saturating half-cycle, so that conduction of the transistor will be maintained by the current gain of the transistor and positive feedback through the magnetic coupling.

2. At the end of the resetting half-cycle, the winding voltage must be made positive enough to make the transistor conduct, either by an overshoot in an underdamped voltage transient, or by the fall-back of the core flux after it is driven to negative saturation.

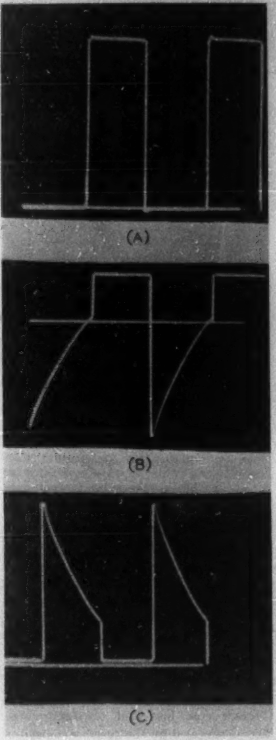


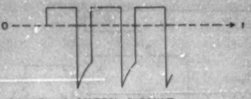
Fig. 7. Oscillograms pertaining to single-transistor magnetic-coupled oscillator with external capacitor (zero lines shown in oscillograms)

A—Load voltage
B—Winding voltage
C—Transistor voltage (emitter to collector)

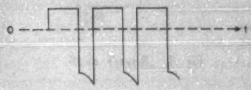
Variety of Circuitry

A variety of circuits may be deduced from the basic single-transistor oscillator circuit in Fig. 4. For instance, the oscillator circuit may be used in a common base connection, as in Fig. 9. It becomes then identical to one half of the oscillator circuit which Uchrin and Taylor of the United States Signal Corps Engineering Laboratories claimed to have developed independently.[4] The germanium p-n-p transistor in Fig. 4 may be replaced by a silicon n-p-n transistor with a slight modification in circuit arrangement, as in Fig. 10. A diode is connected across the emitter and base of the silicon transistor in order to limit the emitter-to-base voltage within the maximum value of 1.5 volts specified by the manufacturer. A constant base drive voltage supplied by the input voltage E_{in} is necessary in this case to start the silicon transistor into conduction.

The output of circuits in Figs. 4, 9, and 10 is a train of d-c rectangular pulses, whose width and frequency may both be varied by adjusting the external capacitor, the input voltage and a current in an additional control winding (not shown in the last two figures). This kind of output is desired and is directly usable in many applications. When an a-c square wave is desired, the d-c output pulses from the single-transistor oscillator may be used to drive a push-pull transistor amplifier as shown in Fig. 11. In this arrangement the diode used in the previous circuits is no longer needed since the base-emitter junction of the transistor $T2$ acts in this capacity. The value of C may be adjusted to obtain any desired degree of symmetry in the a-c square-wave output.



(a) WITH NO CONTROL CURRENT



(b) WITH A LARGE RESETTING CONTROL CURRENT

Fig. 8. Wave forms of winding voltage when core flux is reset to negative saturation

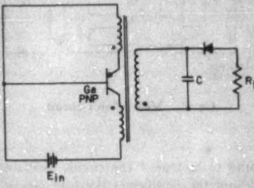


Fig. 9. Single-transistor oscillator circuit in common base connection

Vol. 7
Pt. 1
1956

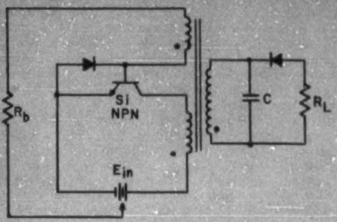


Fig. 10 (left). Single-transistor oscillator circuit using silicon n-p-n transistor

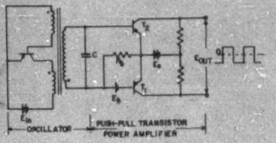
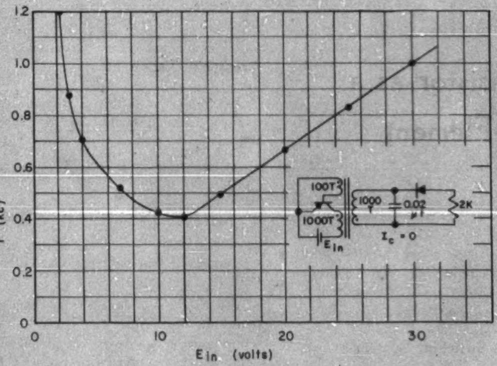


Fig. 11. (left) Converting d-c rectangular pulses to a-c square wave

Fig. 12 (below). Variation of oscillator frequency as function of input voltage



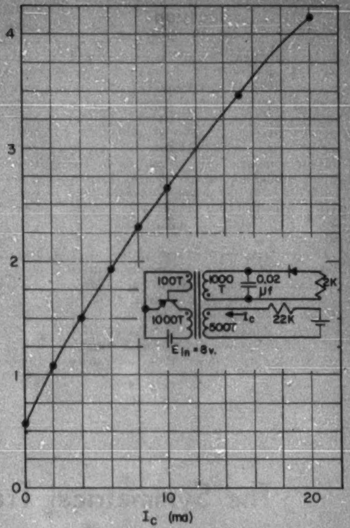


Fig. 13. Variation of oscillator frequency as function of control current

Experimental Results

The variation of frequency of the oscillator as a function of the input voltage, is shown by the curve in Fig. 12. Control of the oscillator frequency within a wide range by means of a positive or saturating control current is illustrated by the curve in Fig. 13. Notice, Fig. 12, that the frequency initially decreases with increasing input voltage. This is explicable since the amount of reset is directly related to the energy stored in the capacitor, which increases as the square of the input voltage, while the time required to drive the core flux from a certain level to saturation decreases only with the first power of the input voltage. The frequency reaches a minimum as the energy stored in the capacitor is just sufficient to reset the core to its negative saturation level. Beyond this point, the frequency increases linearly with the input voltage, as in Royer's circuit.

With a constant input of 8 volts to the circuit of Fig. 4 and no control current, the frequency changes considerably over a temperature range from room temperature to 71 degrees centigrade because the leakage current of the germanium transistor increases substantially at high temperatures, and the leakage current has significant effect on resetting of the core flux. However, the frequency of the circuit in Fig. 10, in which a silicon transistor is used, was found to change only 1/2 per cent over the same temperature range and at the same input voltage. This stability of frequency is attributed to the temperature stability of silicon transistors and the independence of the circuit frequency on the saturation levels of the core flux providing negative saturation is not reached.

Conclusion

A single-transistor magnetic-coupled oscillator has been developed and its principle of operation explained. It is hoped that the detailed circuit analysis has brought forth a fuller understanding of all types of magnetic-coupled oscillators and other similar circuits. The major advantage of the newly developed circuit is its simplicity in that only a single transistor and a single core are required. In addition, the circuit has good stability with temperature variations when silicon transistor is used, and its frequency may be controlled by the input voltage or by a current in a control winding, the latter giving a possible power gain.

References

1. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, G. H. Royer. *AIEE Transactions*, vol. 74, pt. I, July 1955, pp. 322–26.

2. Junction Transistors Used as Switches, R. L. Bright. *Ibid.*, Mar. 1955, pp. 111–21.

3. A Variable Frequency Magnetic-Coupled Multivibrator, R. L. Van Allen. *Ibid.*, vol. 74, pt. I, July 1955, pp. 356–61.

4. A New Self-Excited Square-Wave Transistor Power Oscillator, G. C. Uchrin, W. O. Taylor. Correspondence in *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, no. 1, 1955, p. 99.

(Discussion follows.)

Discussion

R. L. Van Allen (Naval Research Laboratory, Washington, D. C.): In the discussion of the variable frequency magnetic-coupled multivibrator, the paper indicates the effect of stray capacitance in maintaining oscillation, especially at zero control current. It is indeed necessary to consider this capacitance as a prerequisite to sustained oscillation at zero control current. The authors should be congratulated for discerning this fact and for applying this approach to a single-core single-transistor oscillator.

It is interesting that here at the Naval Research Laboratory a practically identical circuit was independently disclosed and investigated by H. T. Mortimer, formerly of the Naval Research Laboratory.[1] Besides describing a single-core single-transistor static converter, Mr. Mortimer's report contains a description and experimental results of a single-core single-transistor d-c to d-c amplifier. The principle of operation as a power amplifier is based on the utilization of an oscillator producing an asymmetric square wave which has a variable fundamental frequency. The waveform of this asymmetric square wave is such that the period of one half-cycle is relatively constant, while the period of the next half-cycle may be controlled as a function of a d-c control voltage. A d-c output is obtained in the form of pulses having constant height and width but with a repetition rate determined by the magnitude of the control voltage.

REFERENCE

1. DC TO AC SQUARE-WAVE STATIC CONVERTER AND AMPLIFIER, H. T. Mortimer. *NRL Report No. 4582*, Naval Research Laboratory, Washington, D. C., July 27, 1955.

Harry T. Mortimer (Hallamore Electronics Company, Anaheim, Calif.): The authors are to be congratulated for their fine presentation of the single-transistor magnetic-coupled oscillator. Their comparison to Mr. Van Allen's circuits as well as the development of their ideas concerning the single transistor oscillator are clear and logical.

It is to be noted, in connection with the work reported here, that the writer had conceived circuitry equivalent in all respects to those reported in this paper. The conception and subsequent development of operating circuits date from November 1954. Development was continued, and led to a patent disclosure in March 1955. The writer's report (see reference 1 of Mr. Van Allen's discussion) fully developed the technical details and contained illustrations showing methods of operation and oscillograph wave forms. This work is considered a part of an unbroken continuation of the magnetic and transistor research program which supported Mr. Van Allen's work.

The Symmetrical Transistor as a Bilateral Switching Element

ROBERT B. TROUSDALE
NONMEMBER AIEE

TRANSISTORS and diodes have been employed successfully as basic building blocks in the design of switching systems. Much progress is being made in the development of transistorized equipment which promises ultimately to supplant the more conventional electromechanical equipment that at present is in common use.

This paper discusses the use of a symmetrical transistor as a bidirectional switch useful in communications systems where very low insertion loss and high ON-OFF ratios are desired. Examples of the use of the switch as a convenient crosspoint as well as a brief analysis of the equivalent circuit and a derivation of circuit impedances leading to optimum performance will be shown.

A Simple Switching System

Fig. 1 is a block diagram of an elementary switching system which shows the use of co-ordinate signal path switch points. These switch points are closed or opened through the action of separate control circuits operating in response to supervisory signals. A simple system such as this can be expanded to larger proportions through the use of more lines and links, and through additional switching stages.

To establish a connection in the system of Fig. 1, it is necessary to close two crosspoints in one of the vertical columns. This completes a signal path between the corresponding two line terminals. A second simultaneous connection can be set up between two different lines through closure of two crosspoints in one of the other vertical columns and, in a similar manner, a third connection can be made through closure of two crosspoints in the remaining vertical column.

Crosspoint Requirements

A crosspoint element suitable for switching service should have the following characteristics:

1. A high degree of electrical isolation between control terminal and controlled terminals.
2. Negligible insertion loss when closed.
3. Negligible signal leakage when open.
4. Negligible load on input and output terminals when open.

Although a simple electromechanical relay possesses these characteristics, a semiconductor crosspoint can claim advantages such as small size and weight, low operating power, high speed, immunity to wear, corrosion, moisture, dirt, shock, and vibration, and, theoretically at least, infinite life.

The Bilateral Transistor Switch

An effective crosspoint can be made through the utilization of a bilateral transistor, as shown in Fig. 2. Bilateral transistors are symmetrical alloy junction germanium units of p-n-p or n-p-n configuration, in which either junction serves equally well as collector or emitter.

Fig. 3 shows the circuit detail of the 3-link system of Fig. 1, in which bilateral transistors are used as crosspoints in an unbalanced connection. The return sides of all lines are connected in common. This arrangement permits the use of a common transformer winding for all transistors in each row as well as a common choke for all transistors in each column.

Paper 56-585, recommended by the AIEE Solid State Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE North Eastern District Meeting, Rochester, N. Y., May 2–4, 1956. Manuscript submitted February 1, 1956; made available for printing April 25, 1956.

ROBERT B. TROUSDALE is with the Stromberg-Carlson Company, Rochester, N. Y.

The co-operation of many people at the Raytheon Manufacturing Company and the General Electric Company in the design and fabrication of bilateral transistors for application in Stromberg-Carlson switching equipment is gratefully acknowledged. Thanks are due to George Elliott, Stromberg-Carlson Company, for his contributions to the work leading to this paper. The research effort reported in this paper was shared by the Navy Bureau of Ships under contract *NObs-61658*.

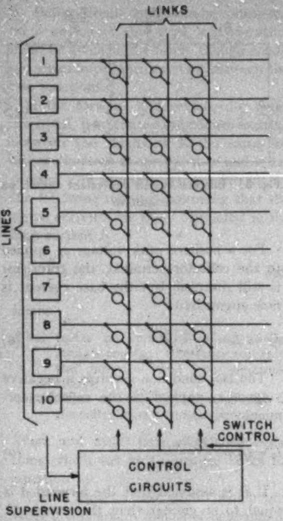


Fig. 1. A 10-line 3-link switching system

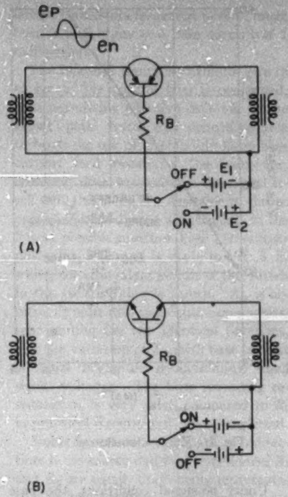


Fig. 2. Basic bilateral switch circuits

A—Shunt-fed p-n-p bilateral transistor switch
B—Shunt-fed n-p-n bilateral transistor switch

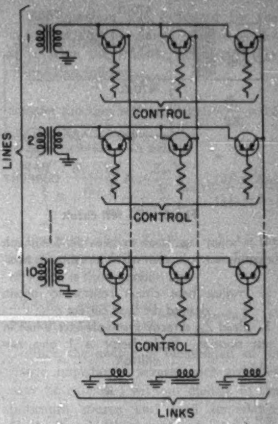


Fig. 3. A 3-link system employing shunt-fed bilateral transistor switches

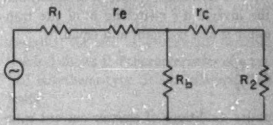


Fig. 4. Bilateral switch equivalent circuit, off condition

It is interesting to note that d-c collector-operating potentials are not employed in the circuit.

Analysis of Bilateral Transistor Switch Operation

Referring again to Fig. 2(A), if it is assumed that the switch is in the OFF condition, then both junctions are biased in the reverse direction. The a-c impedance represented by each junction is very high and equal to the rate of change of the reverse-biasing voltage with respect to the reverse saturation current. The reverse impedances of the junctions are called r_c' and r_e' for the collector and emitter junctions respectively. It should be pointed out that, during the measurement of these reverse-biased junction impedances, the opposite junction is open-circuited.

Transmission in either direction through the switch is attenuated considerably in the OFF condition so long as the positive input signal peaks do not exceed the voltage of the base supply battery E_1. Under these conditions, the a-c equivalent circuit of the switch is as shown in Fig. 4. Typical values are

$R_1 = R_2 = 250$ ohms (source and load impedances)

$R_b = 10{,}000$ ohms

$r_e' = r_c' = 10$ megohms

for which the insertion loss is greater than 140 decibels, neglecting capacitance effects.

It has been found that the impedances of the junctions do not change materially with temperature and, for some germanium transistors tested, impedances measured at 55 degrees centigrade (C) have actually been higher than those at 25 C. When the switch of Fig. 2(A) is in the ON condition, then both junctions are biased in the forward direction, as emitters, and the base current divides about equally between the two. The magnitude of the base current is determined approximately by the ratio between the base supply battery potential E_2 and the base feed resistor R_b.

Assuming transmission from left to right, during positive alternations of the signal, the left-hand junction acts as an emitter and the right-hand junction acts as a saturated collector. Since the voltage drop between collector and emitter terminals of a saturated germanium alloy junction transistor is very small, the collector terminal will follow closely the voltage swings at the emitter terminal.

During negative alternations of the input signal, the left-hand junction acts as a collector and the right-hand junction as an emitter. The transistor will remain in saturation, and the right-hand terminal again closely follows the voltage swings at the left-hand terminal.

In a like manner, positive and negative alternations of a signal originating at the right-hand junction of the transistor are repeated at the left-hand junction. The switch, therefore, can handle signals in either direction and of either polarity, thus giving rise to the name "bilateral."

In developing equations relating transistor parameters and circuit constants to allowable signal levels while the switch is ON, it should be pointed out that the circuit is primarily intended to handle a wide variation in signal levels, including relatively large voltages and currents, and that the transistor is presumed to remain saturated over the entire signal cycle. This being the case, a rigorous analysis of the circuit operation which takes into account variations in transistor parameters as a function of collector or emitter current can become very complex and is assumed to be beyond the scope of this paper. For simplicity, therefore, it is assumed that the d-c switching β of the transistor is a constant and equal to

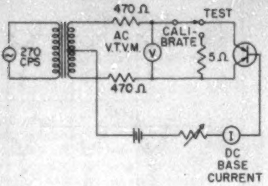


Fig. 5. Z_{ce} test circuit

1. Adjust oscillator to provide 5-millivolt reading on voltmeter when in calibrate position
2. Adjust base circuit resistor to obtain desired d-c base current
3. Read Z_{ce} directly from voltmeter when in test position. Scale factor is 1 ohm per millivolt

$$\beta=\frac{I_c}{I_b} \tag{1}$$

under conditions of collector current saturation. Interchanging the collector and emitter terminals, it may be assumed that the inverse d-c switching β' of the transistor is also constant and equal to

$$\beta'=\frac{I_e}{I_b} \tag{2}$$

under conditions of emitter current saturation. It will also be assumed that the junction voltage drops are small compared to the base supply battery potential E_2 and that the transistor base spreading resistance is a constant r_b. The transformers of Fig. 2(A) are assumed to be ideal.

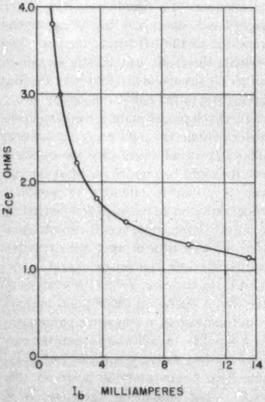


Fig. 6. Z_{ce} as a function of I_b

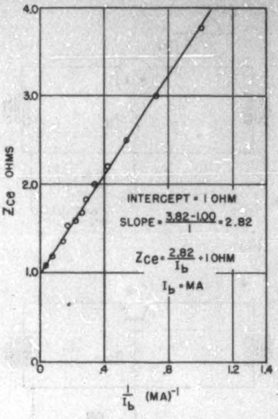


Fig. 7. Z_{ce} as a function of $1/I_b$

Under no-signal conditions, the base current is given by

$$I_b=\frac{E_2}{R_b+r_b} \tag{3}$$

If the signal swing is small compared to E_2, then the transistor will be capable of handling a peak signal current of

$$I_p=\beta I_b=\frac{\beta E_2}{R_b+r_b} \tag{4}$$

if the current flows into the emitter junction, and of

$$I_p=\beta' I_b=\frac{\beta' E_2}{R_b+r_b} \tag{5}$$

if the current flows into the collector junction. At currents in excess of these maxima, the transistor is driven out of its saturated region, and signal clipping or distortion occurs.

If the signal voltage is appreciable in comparison to E_2, as it normally is in practical circuits, equations 4 and 5 must be modified to include the change in base current caused by the signal itself. When a positive peak voltage e_p is applied to the emitter terminal, the base current is increased and the maximum peak signal current becomes

$$I_p=\beta I_b=\frac{\beta(E_2+e_p)}{R_b+r_b} \tag{6}$$

For negative voltages e_n applied to the emitter terminal, the transistor reverses polarity, the base current is reduced, and the peak signal current becomes

$$I_p=\beta' I_b=\frac{\beta'(E_2-e_n)}{R_b+r_b} \tag{7}$$

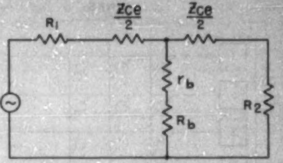


Fig. 8. Bilateral switch equivalent circuit, on condition

For a positive peak voltage e_p applied to the collector terminal, the transistor is still reversed, but the base current is now augmented

$$I_p=\beta' I_b=\frac{\beta'(E_2+e_p)}{R_b+r_b} \tag{8}$$

The last condition involves a negative voltage e_n applied to the collector terminal

$$I_p=\beta I_b=\frac{\beta(E_2-e_n)}{R_b+r_b} \tag{9}$$

If it is assumed that the forward β is equal to or greater than the reverse β', then, in general, equation 7 expresses the smallest value of the peak signal current I_p and is therefore the most limiting of the four equations 6, 7, 8, and 9.

It is desirable to utilize a transistor with high values of β and β' so that the switch may handle a given power level with small base currents and, therefore, high values of R_b. This decreases the shunting effect of R_b across the circuit and decreases the insertion loss. Experimental p-n-p bilateral transistors manufactured by Raytheon and General Electric have exhibited β values in excess of 20 in both directions.

It should also be kept in mind that E_2 must be sufficiently large that the degenerative effect of the input signal on the net base current may be kept within reasonable limits; see equations 7 and

Table I. Characteristics of a Raytheon-Type SC101 Transistor

I_{co} at 6 volts	3.6 microamperes
I_{co} at 25 volts	6.0 microamperes
I_{eo} at 6 volts	2.6 microamperes
I_{eo} at 25 volts	3.8 microamperes
r_c' at 6 volts	6.2×10^6 ohms
r_c' at 25 volts	8.3×10^6 ohms
r_e' at 6 volts	19.6×10^6 ohms
r_e' at 25 volts	3.8×10^6 ohms
Peak inverse voltage, collector junction	81 volts
Peak inverse voltage, emitter junction	93 volts
V_p, punch through	14 volts
Forward $-h_{11}$, 1 milliampere, 5 volts	0.977
Reverse $-h_{11}$, 1 milliampere, 5 volts	0.972
β at $I_c=10$ milliamperes	25.0
β'at $I_e=10$ milliamperes	22.3
Z_{ce} at $I_b=1$ milliampere	3.9 ohms

9. Before circuit design can be completed on a switch of this type, consideration must be given to the finite, though small, a-c series impedance existing between the two junctions.

If the forward and reverse d-c betas are equal, then the a-c impedance existing between the junctions is the same for either direction of current flow and is expressed as a function of base current by the following relation, assuming that the signal currents are small compared to the base current I_b

$$Z_{ce}=\frac{K}{I_b}+r_s \qquad (10)$$

where

K = a factor incorporating β, β', Boltzmann's constant, the temperature of the junction, and the electronic charge

r_s = the sum of the volume resistivities of the collector and emitter junctions

Ebers and Moll[1] have developed equations which act as theoretical bases for equation 10.

Fig. 5 shows a test circuit which provides a convenient means of determining Z_{ce} for symmetrical transistors. In the event that β and β' are unequal, the test circuit indicates the average of the impedance values for the two polarities. Thus, when considering the performance of the bilateral transistor in a network, it is presumed that linear circuit analysis may be performed employing the average Z_{ce} reading, with the understanding that differences between β and β' will give rise to a small amount of harmonic distortion.

Fig. 6 is a curve of Z_{ce} plotted against I_b for a typical Raytheon bilateral transistor. These data were obtained from the test circuit of Fig. 5. Replotting the curve with $1/I_b$ as the abscissa, a straight line is obtained, as shown in Fig. 7. The slope of this curve is the constant K and the intercept is the constant r_s of equation 10. The empirical equation for Z_{ce} of this transistor can then be written as

$$Z_{ce}=\frac{2.82}{I_b}+1.0 \text{ ohm} \qquad (11)$$

where I_b is in milliamperes. This is in approximate agreement with Ebers and Moll,[1] who point out further that the constant K bears an inverse relationship to β and β'. Typical values of Z_{ce} in bilateral transistors measured to date range from 3 to 8 ohms at a base current of 1 milliampere.

The bilateral transistor exhibits gain in so far as the control bias is concerned, thus permitting effective isolation of the signal path from the control source through the use of a relatively large base current feed resistor. The transistor, however, does not amplify the signal itself and, therefore, it is desirable to choose operating impedances which result in the lowest possible insertion loss. To accomplish this, reference is made to Fig. 8, in which the equivalent circuit of the switch in the ON condition is shown. Z_{ce} is arbitrarily split into two equal components, representing the two identical junctions, and the variation of Z_{ce} with base current is made negligible by assuming small signal voltages. The base spreading resistance r_b is very small compared to R_b in practical circuits and can be neglected.

Since the transmission path is bidirectional, the source and load impedances R_1 and R_2 are equal, these being representative of two identical telephones, for example. The transistor can now be looked upon as a 4-terminal network possessing a transfer function which will be closest to unity when the source and load impedances are made equal to the image impedance of the network

$$Z_I=\sqrt{Z_{oc}Z_{sc}}$$

$$=\sqrt{\left(\frac{Z_{ce}}{2}+R_b\right)\left[\frac{Z_{ce}}{2}+\frac{Z_{ce}R_b}{2\left(\frac{Z_{ce}}{2}+R_b\right)}\right]}$$

$$=\sqrt{\frac{Z_{ce}^2+4R_bZ_{ce}}{4}}$$

Since $R_b>>Z_{ce}$ in most practical circuits, this reduces to

$$Z_I=\sqrt{R_bZ_{ce}} \qquad (12)$$

If, for example, a 10,000-ohm R_b is employed with a transistor exhibiting a Z_{ce} of 5 ohms, then the source and load impedances should be made equal to

$$R_1=R_2=\sqrt{10{,}000\times 5}=224 \text{ ohms}$$

This impedance may be realized by the proper choice of turns-ratio for the line transformers in Fig. 3. The link chokes serve merely as d-c returns and do not exhibit appreciable a-c shunt loss.

The insertion loss exhibited by the switch of the foregoing example is derived from the transfer function of the network as follows:

$$\text{transfer function without switch}=\frac{R_1}{2R_1}=0.5$$

transfer function with switch

$$=\left(\frac{R_1}{R_1+\frac{Z_{ce}}{2}}\right)\left(\frac{R_1-\frac{Z_{ce}}{2}}{2R_1}\right)=0.5\frac{(2R_1-Z_{ce})}{(2R_1+Z_{ce})}$$

The factor $(2R_1-Z_{ce})/(2R_1+Z_{ce})$ is the insertion gain of the switch, the reciprocal of which is the insertion loss

$$\frac{2R_1+Z_{ce}}{2R_1-Z_{ce}}=\frac{448+5}{448-5}=\frac{453}{443}=0.194 \text{ decibel} \qquad (13)$$

Since the switch is terminated in its image impedance, several such switches may be cascaded without change in the optimum source and load impedance levels, the resulting insertion loss being determined by multiplying the decibel insertion loss per stage by the number of stages.

To bring out more fully the unique characteristics of transistors specifically designed for use as crosspoint switches, Table I shows the characteristics of a typical Raytheon-type *SC101* bilateral transistor.

Summary

The symmetrical germanium alloy junction transistor has been found to be a very satisfactory bidirectional switch for use in electronic switching systems. It provides circuit isolation between the signal path and the control source, and can handle signals of either polarity. With the switch open, signal leakage is down more than 100 decibels even at high temperatures, and insertion losses of less than 0.2 decibel can be obtained with the switch closed. The degree of circuit isolation, the insertion loss, and the power-handling capacity can all be improved through the use of transistors with high forward and reverse current gains.

Reference

1. LARGE-SIGNAL BEHAVIOR OF JUNCTION TRANSISTORS, J. J. Ebers, J. L. Moll. *Proceedings, Institute of Radio Engineers*, New York, N. Y., Dec. 1954, pp. 1761–72.

An Analysis and Design of the Ferroelectric Resonant Trigger Pair

CHARLES E. GREMER
NONMEMBER AIEE

THE CIRCUITS to be analyzed in this paper are the 2-branch series-resonant circuit shown in Fig. 1; and the 2-mesh, parallel resonant circuit shown in Fig. 2. There are several approaches for analyzing such circuits based upon developments by Duffing, Poincaré, Kryloff and Bogoliuboff and others.[1-9] The choice of one approach over another depends on the accuracy desired, the accuracy obtainable, and the exact problem confronting the analyst. It should be noted that all of the known analytical methods suffer from one inherent disadvantage: they produce solutions for only the case under consideration without any hint as to the totality of solutions available. A qualitative method, known as the topological approach, gives an over-all picture of the totality of solutions but yields no specific information on any particular case.[3,5,6]

In any circuit containing nonlinearities, it must be decided what function will represent the nonlinearity before the problem is commenced. Since this paper concerns itself with the nonlinear capacitor as its vehicle of operation, a charge function of the voltage across the capacitor terminals is desired. Now it is known a priori that two states are to exist for the branches: 1. a large current condition, and 2. a small current condition. Since it is desired to obtain the criteria for such states, and since these two states depend on whether or not the branch is in nonlinear resonance, the primary interest in the charge-voltage relationship is not their actual point-by-point correspondence, but rather an average of the initial and an average of the final charge-voltage slopes of the loop tip locus.

Now since a single-valued function is to be specified and it is actually known that the relationship is double-valued for any given driving voltage, accuracy will be forfeited *ab initio*. It is thus fortuitous to choose a rather simple function which still defines, within limits, the circuit operation.

The function chosen is

$$e(q)=a_1q+a_3q^3 \qquad (1)$$

where a_1 and a_3 are constants to be determined by measurements on the nonlinear capacitor. Obviously, any odd function which approximates the $e(q)$ relation could be used. However, owing to the initial assumptions, it is believed that nothing greater than a cubic is warranted.

Two-Branch Series-Resonant Circuit

THE LOSS-FREE CIRCUIT

The circuit equations for Fig. 1 can be written

$$q_s=q_1+q_2 \qquad (2)$$

$$L(\ddot{q}_1-\ddot{q}_2)+e_1-e_2=0 \qquad (3)$$

Using equation 1, equation 3 becomes

$$(\ddot{q}_1-\ddot{q}_2)+\frac{a_1}{L}(q_1-q_2)+\frac{a_3}{L}(q_1^3-q_2^3)=0 \qquad (4)$$

At this point a mathematical artifice is introduced. Equation 4 is first solved for the difference differential and a term ω^2-(q_1-q_2) is then added to both sides of this equation. This procedure results in the final form of the circuit equation

$$(\ddot{q}_1-\ddot{q}_2)+\omega^2(q_1-q_2) =\left(\omega^2-\frac{a_1}{L}\right)(q_1-q_2)-\frac{a_3}{L}(q_1^3-q_2^3) \qquad (5)$$

Now, the procedure is one of iteration, where an approximation is substituted into the right-hand side of equation 5 and the left-hand side is solved for the steady-state periodic solution.

The first approximation begins with

$$q_{10}=A \sin \omega t \text{ and} \qquad (6)$$
$$q_{20}=B \sin \omega t$$

where the second subscript denotes the iteration by which the value was obtained.

Equation 5 thus becomes

$$(\ddot{q}_1-\ddot{q}_2)+\omega^2(q_1-q_2) =\left[\left(\omega^2-\frac{a_1}{L}\right)(A-B)-\frac{3a_3}{4L}\times (A^3-B^3)\right] \sin \omega t+ \frac{a_3}{4L}(A^3-B^3) \sin 3\omega t \qquad (7)$$

Since $a_3>0$, only periodic solutions will be available.[6] Invoking the condition for periodicity (i.e., due to the $\omega^2(q_1-q_2)$ term, the coefficient of sin ωt must vanish or else secular terms will appear in the solution), there results

$$\left(\omega^2-\frac{a_1}{L}\right)(A-B)-\frac{3a_3}{4L}(A^3-B^3)=0 \qquad (8)$$

Disregarding the trivial case, $A=B$, and from equation 2 there result two equations relating A and B. These are the following

$$\psi_1\equiv 0=-Q_s+A+B \qquad (9)$$

$$\psi_2\equiv 0=-S+m(A^2+AB+B^2) \qquad (10)$$

where

$$S=(\omega/\omega_0)^2-1 \qquad (11)$$

$$\omega_0^2=a_1/L \qquad (12)$$

$$m=3a_3/4a_1 \qquad (13)$$

The only information available from ψ_1 and ψ_2 is that for ψ_2 to be a real ellipse $S\overset{M}{>}0$, and hence $\omega\overset{M}{>}\omega_0$, where the symbol $\overset{M}{>}$ means "must be greater than."

After the condition for periodicity has been invoked, equation 7 is then solved and there is thus obtained

$$(q_1-q_2)_1=K \sin \omega t- \frac{a_3}{32\omega^2L}(A^3-B^3) \sin 3\omega t \qquad (14)$$

It is now assumed that the constant of integration K is equal to the factor $A-B$. There is thus obtained by equating the portions of this equation identifiable with q_1 and q_2 respectively

$$q_{11}=A \sin \omega t-DA^3 \sin 3\omega t \qquad (15A)$$

$$q_{21}=B \sin \omega t-DB^3 \sin 3\omega t \qquad (15B)$$

where

$$D=\frac{a_3}{32\omega^2L}=\frac{m}{24(S+1)} \qquad (16)$$

These relationships are now substituted into the right-hand side of equation 5 and the condition for periodicity again invoked. Disregarding the trivial solution, $A=B$, there results

---

Paper 56-119, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956; and re-presented at the North Eastern District Meeting, Rochester, N. Y., May 2–4, 1956. Manuscript submitted June 27, 1955; made available for printing March 14, 1956.

LT. CHARLES E. GREMER is at the U. S. Naval Shipyard, Charleston, S. C.

The author wishes to thank Dr. Charles F. Spitzer, of the General Electric Company, who first suggested work on the problem presented, and who discussed the solutions. Gratitude is also expressed to the General Electric Company, Syracuse, N. Y., for providing laboratory space and equipment. The project was completed as a partial requirement for a United States Naval Postgraduate School M.S. degree, Monterey, Calif. In this connection thanks are due Prof. Carl E. Menneken and Capt. Paul Van Leunen of the Naval Postgraduate School; also C. P. Gade and C. A. Rosen of the General Electric Company for arranging the industrial tour where the project was accomplished.

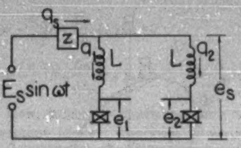


Fig. 1. Two-branch nonlinear series-resonant circuit

$$\psi_3 \equiv 0 = -S + m(A^2+AB+B^2) + mD(A^4+A^3B+A^2B^2+AB^3+B^4) + 2mD^2(A^6+A^5B+A^4B^2+A^3B^3+A^2B^4+AB^5+B^6) \quad (17)$$

To simplify the solution of equation 17 and to determine more easily whether or not it has only one real root, the rotation matrix

$$\begin{Vmatrix} A \\ B \end{Vmatrix} = \begin{Vmatrix} \sqrt{\tfrac{1}{2}} & -\sqrt{\tfrac{1}{2}} \\ \sqrt{\tfrac{1}{2}} & \sqrt{\tfrac{1}{2}} \end{Vmatrix} \times \begin{Vmatrix} \alpha \\ \beta \end{Vmatrix} \quad (18)$$

and the linear transformation

$$\alpha = \rho \cos \theta \quad (19A)$$

$$\beta = \rho \sin \theta \quad (19B)$$

are applied to equation 17, which results in

$$\psi_3 = 0 = -S/m + {}^1/_2 f(\theta)\rho^2 + {}^1/_4 Dg(\theta)\rho^4 + {}^1/_4 D^2 h(\theta)\rho^1 \quad (20)$$

where

$$f(\theta) = 2 + \cos 2\theta \quad (21)$$

$$g(\theta) = 3.5 + 2 \cos 2\theta - {}^1/_2 \cos 4\theta \quad (22)$$

$$h(\theta) = 15 + \cos 2\theta[4 - 18 \cos 2\theta - \cos^2 2\theta] + 7 \cos 4\theta \quad (23)$$

For any θ, ψ_3 has only two finite, real roots (positive and negative values of the same magnitude) and hence represents a single, closed loop. Thus equation 20 need be solved for only one value of ρ^2, and hence ρ, for each θ. ψ_1, ψ_2, and ψ_3 are shown in Fig. 3.

From Fig. 3 the mode of operation may be explained. For small values of E_s, and hence Q_s, the two series branches behave as linear networks and thus A equals B. As E_s is increased, ψ_1 moves out from the origin, remaining parallel to the β axis. The operating point remains on the intersection of ψ_1 and the α axis. It can be shown that ψ_1 will be tangent to ψ_3 at P_1. When ψ_1 does become tangent to ψ_3 by increased Q_s, two other points of operation are available: P_2 and P_3. Although no mathematical proof is given here, it can be deduced from the work of Stoker that operation in the area around P_1 is unstable.[6]

The value of source current required to reach bistability condition can be solved from equation 20 and equation 2. It is given by

$$I_s = \omega Q_s \quad (24)$$

$$Q_s = \sqrt{2}\rho(\theta = 0°) \quad (25)$$

$$\rho(\theta = 0°) = \frac{0.756}{\sqrt{D}}(1.68 \sinh u - 0.417)^{1/2} \quad (26)$$

$$u = \frac{1}{3} \sinh^{-1} \frac{0.915S + 0.805}{S+1} \quad (27)$$

Obviously, if ψ_1 moves out so far that it fails to intersect ψ_3, bistability fails to exist and both branches revert to equal but large currents. This value can be obtained by finding the appropriate point besides P_1 where $d\alpha/d\beta$ vanishes. In view of the original approximations, this current can be obtained to a sufficient degree of accuracy from a plot such as Fig. 3.

Low Loss Circuit

Attention is now given to the case where the two branches contain low losses. Equation 5 is rewritten to include the effect of resistance as follows

$$(\ddot{q}_1 - \ddot{q}_2) + \omega^2(q_1 - q_2) = \left(\omega^2 - \frac{a_1}{L}\right)(q_1 - q_2) - \frac{R}{L}(\dot{q}_1 - \dot{q}_2) + \frac{a_3}{L}(q_1^3 - q_2^3) \quad (28)$$

Assuming that R is small compared to Z_1, then I_s will still be primarily determined by Z_1. For the remainder of the 2-branch series-resonant discussion, it will be assumed that $A \gg B$. The first iteration begins with the approximations

$$q_{10} = A \sin \omega t \quad (29A)$$

$$q_{20} = B \sin (\omega t - \phi) \quad (29B)$$

Using these values in the right-hand side of equation 28 and application of the periodicity conditions results in the modified ellipse ψ_2' where it has been assumed that $S \gg mB^2$

$$\psi_2' \equiv 0 = A^2 + AB + B^2 - S/m + \frac{2_R2}{a_1^2 Sm} \quad (30)$$

Now, ψ_2' will not be a real ellipse and hence bistability cannot result unless

$$R < a_1 S/\omega \quad (31)$$

Under the assumption that $B \approx 0$, the following relations can be derived

$$A = \sqrt{\frac{S}{m} - \frac{\omega^2 R^2}{a_1^2 Sm}} \quad (32)$$

$$B = \frac{\omega R}{a_1 S} A \quad (33)$$

$$\Delta = A/B = \frac{a_1 S}{\omega R} \quad (34)$$

$$E_1 = a_1 A + a_3 A^3 \quad (35)$$

$$E_2 = a_1 B \quad (36)$$

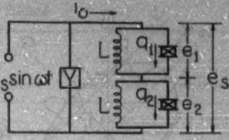


Fig. 2. Two-mesh nonlinear parallel resonant circuit

From these equations two important relations between the sampling voltages, E_1 and E_2, can be obtained. These are the difference between the two sampled voltages and the ratio of the two sampled voltages. This difference determines the quality of the bistable circuit, or at least the ability to distinguish between the two stable states. With a little mathematical manipulation there is obtained

$$E_1 - E_2 = \left[a_1(1+S) - \frac{\omega R}{S}\left(\frac{\omega R}{a_1} - 1\right)\right] \times \sqrt{\frac{S}{m} - \frac{\omega^2 R^2}{a_1^2 Sm}} \quad (37)$$

$$E_1/E_2 = \Delta\left(1 + S - \frac{S}{\Delta^2}\right) \quad (38)$$

The smaller the value of m is the larger the potential difference between the nonlinear capacitors is for a given Δ. This follows from equation 37 and the fact that Δ is related to the parameter S by equation 34. This potential difference, however, is obtainable only by making A increase its value over B. But since the switching ratio Δ was prescribed, both A and B increase. Although Δ, itself, is independent of m, it must be remembered that too small a nonlinearity results in currents of such magnitude that the components may be destroyed before the bistability conditions can be reached.

The second iteration procedure modifies these relations somewhat and emphasizes the choice of m. It can be shown that the nonlinearity should be such that

$$\frac{S}{S - mB^4} \geq \Delta > 1 \quad (39)$$

does not require B so large that the component ratings are exceeded by the value of A.

Two-Mesh, Parallel Resonant Circuit

The 2-mesh, parallel resonant case can now be treated with less detail. From Fig. 2 the following equations can be written

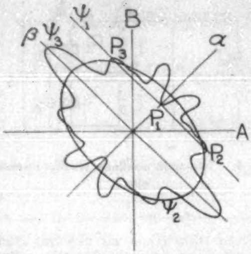


Fig. 3. First and second iteration solution curves for series-resonant circuit

$$i_0=\frac{1}{L}\int e_1dt+\dot{q}_1 \tag{40}$$

$$i_0=\frac{1}{L}\int e_2dt+\dot{q}_2 \tag{41}$$

Subtracting equation 41 from equation 40 there results

$$\frac{1}{L}\int(e_1-e_2)dt+(\dot{q}_1-\dot{q}_2)=0 \tag{42}$$

Since there are no d-c sources in the circuit, no loss of generality occurs if equation 42 is differentiated with respect to time. This results in

$$(\ddot{q}_1-\ddot{q}_2)+\frac{1}{L}(e_1-e_2)=0 \tag{43}$$

The other basic equation is

$$e_1+e_2=e_s \tag{44}$$

In lieu of the constant current source, a constant voltage source is prescribed, and hence i_0 is a small quantity. For lossless circuits, e_1 and e_2 both lag i_0 by 90 degrees if $\omega>\omega_0$. The charge will be in phase with the voltage, and the first approximations can be written as

$$q_{10}=A\cos\omega t \tag{45A}$$

$$q_{20}=B\cos\omega t \tag{45B}$$

Proceeding as with the series-resonant case, the two solution curves relating A and B are found to be

$$\psi_2\equiv 0=-S+m(A^2+AB+B^2) \tag{46}$$

$$\psi_4\equiv 0=-\frac{E_s}{a_1}+(A+B)+m(A^3+B^3) \tag{47}$$

It is to be noted that ψ_2 is common to both systems, but that the straight line ψ_1 has now become the cubic ψ_4. The fact that ψ_4 is a cubic, as shown in Fig. 4, makes it possible to demonstrate bistability from the first iteration. The operating point for low values of E_s is on the intersection of ψ_4 and the line $A=B$. As E_s is increased, ψ_4 moves away from the origin, its asymptotic line remaining parallel to the line $A=-B$. When ψ_4 becomes tangent to ψ_2 at P_1, the operating point "jumps" from P_1 to either P_2 or P_3.

The half-axes of the ellipse are

$$a_\alpha=\sqrt{\frac{2S}{3m}} \tag{48}$$

$$a_\beta=\sqrt{\frac{2S}{m}} \tag{49}$$

The first of these represents the "distance" from the origin to P_1. Since $A=B$ at this point

$$A=\sqrt{\frac{S}{3m}} \tag{50}$$

Substituting this value in equation 47 results in

$$E_s=2a_1\left(1+\frac{S}{3}\right)\sqrt{\frac{S}{3m}} \tag{51}$$

as the source voltage required for bistability. This value of E_s has another interpretation which may be obtained from a study of Fig. 4. If E_s were larger than this value, then there is the possibility of having two sets of bistable operating points. Thus for no more than two stable operating points

$$E_s\leq 2a_1\left(1+\frac{S}{3}\right)\sqrt{\frac{S}{3m}} \tag{52}$$

As E_s is further increased, there is a value beyond which bistability will cease and reversion to the $A=B$ condition occurs. This value of E_s is found by determining the common point of tangency between ψ_2 and ψ_4 besides the one already found at P_1. For ψ_2 and ψ_4 to osculate at a given point, the following conditions must be satisfied at the point

$$\frac{\partial\psi_2}{\partial A}=\frac{\partial\psi_4}{\partial A} \tag{53A}$$

$$\frac{\partial\psi_2}{\partial B}=\frac{\partial\psi_4}{\partial B} \tag{53B}$$

These give

$$1+3mA^2=m(2A+B) \tag{54}$$

$$1+3mB^2=m(2B+A) \tag{55}$$

Division of these two equations, simplification and omission of the $A=B$ solution results in

$$\psi_5\equiv 0=A^2+3AB+B^2-\frac{1}{3m} \tag{56}$$

Application of the rotation matrix, equation 18 produces a hyperbola

$$\psi_6\equiv 0=5_\alpha{}^2-\beta^2-\frac{2}{3m} \tag{57}$$

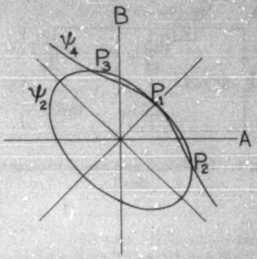


Fig. 4. First iteration solution curves for parallel resonant circuit

with half-axes

$$b_\alpha=\sqrt{\frac{2}{15m}} \tag{58}$$

$$b_\beta=\sqrt{\frac{2}{3m}} \tag{59}$$

The ellipse ψ_2 and the hyperbola ψ_6 will intersect if the half-axes obey the relations

$$b_\alpha<a_\alpha,\text{ or }\sqrt{\frac{2}{15m}}<\sqrt{\frac{2S}{3m}} \tag{60}$$

From this inequality another criterion for bistability results

$$S>1/5,\text{ or} \tag{61}$$

$$\omega^2/6\omega_0{}^2/5 \tag{62}$$

A second iteration can be considered and accomplished as for the series-resonant circuit: likewise, the low loss case can be treated, all of which has been done, but the results are omitted here in order to present another method of obtaining the bistability criteria.[10]

A Second Method for Bistable Current

Instead of writing equation 3, the branch equations can be written as

$$e_0=L\dot{q}_1+e_1 \tag{63}$$

$$e_0=L\dot{q}_2+e_2 \tag{64}$$

where e_0 is as shown in Fig. 1 and is also given by

$$e_0=e_s-e_z(q_1+q_2) \tag{65}$$

the term $e_z(q_1+q_2)$ being a linear function representing the voltage drop across Z. From the first iteration assumptions, e_0 is of the form $E_0\sin t$ to a first approximation. Hence, upon modifying equation 63 and equation 64 as was equation 5; and after application of the periodicity conditions there is obtained

$$\left(\omega^2-\frac{a_1}{L}\right)A-\frac{3a_3}{4L}A^3+E_0=0 \tag{66}$$

$$\left(\omega^2-\frac{a_1}{L}\right)B-\frac{3a_3}{4L}B^3+E_0=0 \tag{67}$$

Now, Stoker has shown that for a first iteration, at least, the locus of vertical tangents of the curves defined by equation 66 or equation 67 are unstable.[6] The locus is found by letting $d\omega/dA$ (or $d\omega/dB$) vanish. The result is, for equation 66, the following

$$S-3mA^2=0 \tag{68}$$

from which

$$A=\sqrt{\frac{S}{3m}} \tag{69}$$

which is the same value as given by equation 50.

Thus the first iteration value of required bistability current can be obtained by consideration of one branch alone. The same procedure holds true for the second iteration. Application of the periodicity conditions to this approximation gives

$$SA-mA^3-mDA^5-2mD^2A^7+\frac{L}{a_1}E_0=0 \tag{70}$$

Letting $d\omega/dA$ vanish produces the following result

$$\frac{S}{m}-3A^2-5DA^4-14D^2A^6=0 \tag{71}$$

If the value $B=A$ is set into equation 17, the same result is obtained. Thus if the hypothesis is accepted that the locus of vertical tangents is also a locus of unstable points, then the reasoning that leads to equation 24 through equation 27 is correct. Conversely, the required value of current (or voltage) can be obtained by considering one branch alone. However, such items as switching ratio and loss limitations are not readily available.

Conclusions

Special note should be made of the fact that the switching ratio is independent of the degree of nonlinearity. However, too small a nonlinearity may result in currents which cause the components to be destroyed. On the other hand, too large a nonlinearity could cause the output voltages to be too small for use. Experimental evidence confirms the validity of equation 24 through equation 27 within the accuracy of component values.[10] Table I gives a brief résumé of these results. The component values were measured at 20 kc, the highest frequency at which measurements could be made on the nonlinear capacitors during the time alloted to the experiments.

Table I. Validity Résumé

				Constants	
f, kc	S	I_s Component	I_s Exponent	a_1, Volts per Coulomb	a_3, Volts per Coulomb^3
20	0.94	73 ma*	70 ma	9×10^8	9.16×10^{21}
233	2.86	159 amperes	0.544 amperes		
233	0.091	484 ma.	327 ma		

*Milliamperes.

References

1. Erzwungene Schwingungen bei veränderlicher Eigenfrequenz, G. Duffing. F. Vieweg u. Sohn, Braunschweig, Germany, 1918.

2. Les méthodes nouvelles de la céleste, H. Poincare. Gauthier-Villars, Paris, France, 1892.

3. Some Analytical Methods of Solving Nonlinear Differential Equations, C. E. Gremer. *Technical Information Series No. R55ELP25*, General Electric Company, Schenectady, N. Y., Feb. 7, 1955.

4. Introduction to Nonlinear Mechanics (translation by S. Lefschetz), N. Kryloff, N. Boboliuboff. Princeton University Press, Princeton, N. J., 1947.

5. Non-Linear Mechanics (book), N. Minorsky. J. W. Edwards, Ann Arbor, Mich., 1947.

6. Nonlinear Vibrations (book), J. J. Stoker. Interscience Publishers, Inc., New York, N. Y., 1950.

7. Applied Mathematics for Engineers and Physicists (book), L. A. Pipes. McGraw-Hill Book Company, Inc., New York, N. Y., 1947.

8. Applications of Integral Equations to the Solution of Nonlinear Electric Circuit Problems, L. A. Pipes. *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 445–50.

9. A Mathematical Analysis of a Series Circuit Containing a Nonlinear Capacitor, Louis A. Pipes. *Ibid.*, July 1953, pp. 238–44.

10. Analysis and Design of the Ferroelectric Resonant Trigger Pair, C. E. Gremer. *M.S. Thesis*, U. S. Naval Postgraduate School, Monterey, Calif., 1955.

11. The Ferroresonant Trigger Pair: Analysis and Design, C. F. Spitzer. *Technical Information Series No. 54-E-292*, General Electric Company, Schenectady, N. Y., Oct. 27, 1954.

The Ferroresonant Trigger Pair: Analysis and Design

CHARLES F. SPITZER
MEMBER AIEE

Synopsis: The phenomenon of nonlinear resonance can be utilized in the design of a trigger pair which uses neither vacuum tubes nor transistors, is capable of high-speed operation, and operates with electrical input and output signals. This article outlines the analysis of the basic circuit to a first approximation only, and design relations are derived. Experimental data are also given. Applications of the circuit described may be found in the logic components of digital data-handling apparatus.

IF, IN A LOW-LOSS series resonant circuit, Fig. 1(A), one or both energy storage elements are nonlinear, and the system is supplied from an a-c power source of suitable voltage and frequency, then the effective value of the resultant nonsinusoidal current may have either of two possible values; see Fig. 1(B). Thus, basically, this circuit can be used as a single-ended bistable element: the output is either high or low.

If two such nonlinear circuits are connected in parallel, and the system is supplied from a power source of suitable current and frequency, then the source cur-

---

Paper 56-584, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE North Eastern District Meeting, Rochester, N. Y., May 2–4, 1956. Manuscript submitted July 1, 1955; made available for printing February 24, 1956.

Charles F. Spitzer is with General Electric Company, Syracuse, N. Y.

This paper is condensed from a dissertation on ferroresonant circuits for digital computers, presented to the Yale University School of Engineering in partial fulfillment of the requirements for the degree of doctor of engineering. The work was carried out at the Electronics Laboratory of the General Electric Company's electronics division, Syracuse, N. Y.

The author expresses sincere appreciation to Dr. L. C. Meier and Charles A. Rosen for their permission to utilize facilities of the electronics laboratory for some of the work, and to Dr. John M. Blank for supplying the ferrite cores used in the experiments. The continued interest and advice given by Prof. P. F. Ordung of Yale University are gratefully acknowledged.

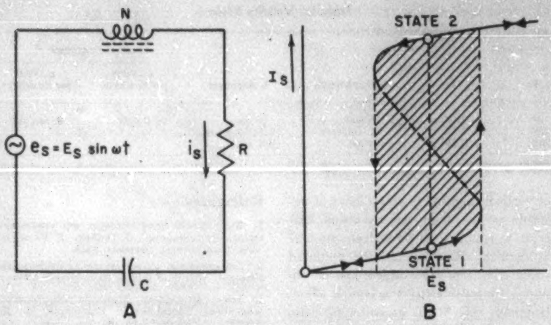


Fig. 1(A)—Basic nonlinearly resonant series circuit, with saturable core. (B)—Response of the basic nonlinearly resonant series circuit of (A)

rent flows almost entirely in either one branch or the other. Thus, this circuit can be used as a double-ended bistable element: the output of one branch is high, when the output of the other branch is low, and vice versa; see Figs. 2 and 3. Triggering can be achieved in a variety of ways, either by means of electrical signals (current pulses through trigger windings) or magnetic signals directly introduced into one or the other core, or into both cores simultaneously.

Suitably designed, the circuit of Fig. 2 can thus be used as the basic unit for binary counters, for storage registers, for gates, and for a variety of other logic components for digital data-handling systems. Triggering for this application is usually achieved most conveniently by successive like-polarity pulses, applied either to a common trigger coil surrounding the nonlinear cores of the two branches, or to the terminals of a series connection or parallel connection of two trigger coils, each of which surrounds the nonlinear core of one of the two branches.

The manner of triggering and the power requirements to achieve it are not the subject of this paper. Rather, it is desired to determine those parameter relations of the basic double-branch circuit, which will maximize the ratio of the branch currents, and therefore result in optimum design for the applications just listed.

The basic ferroresonant trigger pair is shown schematically in Fig. 2, and consists of a parallel arrangement of two series circuits, each containing a capacitor and a nonlinear inductor. The power source is a constant-voltage generator in series with a suitable impedance, such as capacitor C_2. This circuit appears to have been independently developed by Isborn,[1] and by this author and Robert J. Reich.[2]

The underlying phenomenon of nonlinear resonance was reported as early as 45 years ago,[3] and has been the object of study for many years.[4-6] A survey of the literature on nonlinear resonance discloses that the exact mathematical analysis of even the basic 1-mesh nonlinear series circuit is relatively involved.[7-10] The more complicated 2-mesh nonlinear device appears discouraging, from the viewpoint of rigorous analysis. From a practical viewpoint, however, one must realize that the first premise, i.e., that of a simple mathematical description of the nonlinearities in the circuit, seriously and a priori abrogates the accuracy of the result, and no amount of rigor in subsequent steps can compensate for the errors introduced by the initial assumptions. It is therefore the purpose of this paper to outline a simplified analysis of the double-branch series-ferroresonant trigger circuit, to a first approximation only. The fundamental components of voltages and currents are considered, and all harmonics are disregarded, in the evolution of design equations of sufficient simplicity and accuracy to be useful to the engineer constructing a trigger pair. It is believed that this initial objective has been reached. If desired, however, better accuracy can be achieved by use of the well-known iteration method,[7,8] for which the subsequent results represent the first approximation.

The assumption that the fundamental component is sufficiently predominant to permit the exclusion of all harmonics in this analysis can perhaps be made almost acceptable. It will be seen later that a small degree of nonlinearity causes the output voltage to become refreshingly large, even though this is accompanied by relatively large currents throughout the system. This does not imply that the harmonics of all currents and all voltages are always negligible, as is readily seen in the oscillograms of Fig. 4 for example, but the small degree of nonlinearity nevertheless lends some small measure of justification to the procedure.

Nomenclature

All units are in the rationalized meter-kilogram-second system.

a_1, a_3 = coefficients in the flux-current relationship
a_α, a_β = major and minor axes of operating ellipse
b_α, b_β = major and minor axes of limiting hyperbola
A = peak value of flux in core of coil carrying larger of two branch currents
B = peak value of flux in core of coil carrying smaller of two branch currents
C_1 = capacitance in series with a-c source
C_2 = capacitance in each of two parallel branches
e_0, E_0 = instantaneous and peak values of branch voltage
e_1, E_1 = instantaneous and peak values of larger of two output voltages
e_2, E_2 = instantaneous and peak values of smaller of two output voltages
e_s, E_s = instantaneous and peak values of a-c source voltage
i_1, I_1 = instantaneous and peak values of larger of two branch currents
i_2, I_2 = instantaneous and peak values of smaller of two branch currents
i_s, I_s = instantaneous and peak values of a-c source current
K_1, K_2 = specific values for A and B respectively
L_0 = unsaturated coil inductance
L_i = differential coil inductance, at coil current i
m = symbol for ratio $3a_3/4a_1$
N = number of turns on each load coil
P_1, P_2 = operating points, defined by intersection of ψ_1 and ψ_2
R = resistance of each of two branch circuits
S = symbol for term $(\omega^2/\omega_0^2)-1$
t = time variable
Z_2 = unsaturated impedance of branch carrying smaller of two branch currents
$\approx$ = symbol denoting an approximate equality
$\triangleq$ = symbol denoting a definition
α, β = co-ordinates resulting from transformation of A,B co-ordinates
δ = phase angle between e_s and the time derivative of e_0
Δ = switching ratio, $=I_1/I_2$
ω_0 = unsaturated resonant angular frequency of either branch
ω = angular frequency of source
θ = phase angle between I_1 and I_2
φ_1 = instantaneous flux in the core of the coil carrying larger of two branch currents
φ_2 = instantaneous flux in core of coil carrying smaller of two branch currents
ψ_1 = symbol designating operating ellipse
ψ_2 = symbol designating intersecting curve
ψ_3 = symbol designating limiting hyperbola

In using the final design equations, it

Vol. 7
Pt. 1
1956

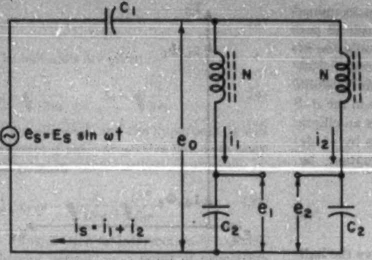
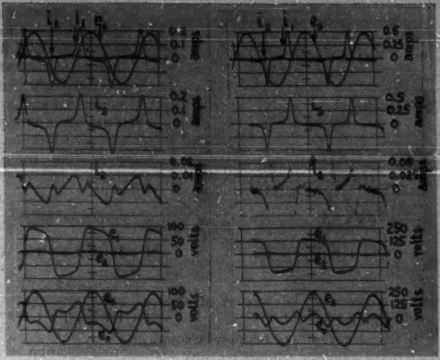


Fig. 2 (above). Schematic diagram of the loss-free ferroresonant trigger pair

Fig. 3 (right). Wave shapes of all currents and voltages for $E_s = 100$ volts, and for $E_s = 250$ volts. Frequency: $f = 233$ kc



must be carefully recognized that all expressions for voltages and currents in terms of the circuit parameters are given as peak values and not as effective or rms values.

Section I. Limiting Parameter Values

Consider the system shown in Fig. 2. If this circuit is to be bistable, certain necessary relations will need to exist between the parameters. In this section some of these basic relations are explored considering the system to be loss-free.

Since the two coils are identical, and since the flux-current relationship is assumed to be nonlinear, it follows that, ignoring all hysteresis effects, the currents can be related to their respective core fluxes by power series with identical coefficients. However, inasmuch as the calculations become staggeringly complicated with the use of terms of higher order, and in view of the quesionable engineering value of the increased accuracy resulting from the use of higher-order terms, the following calculations will not consider terms beyond the third power. Thus, the following relations are assumed

$$i_1 = a_1\varphi_1 + a_3\varphi_1^3 \qquad i_2 = a_1\varphi_2 + a_3\varphi_2^3 \tag{1}$$

where i_1 and i_2 are the branch currents, φ_1 and φ_2 are the corresponding fluxes, and a_1 and a_3 are constants defined by the core material, the core geometry, and the number of turns on each core.

Little is gained in a practical circuit, with the introduction of even the fifth-power term, which case has been treated in some detail by this author, but is here omitted to conserve space.[11,12]

Two equations can be written for the circuit

$$i_s = i_1 + i_2 \tag{2}$$

$$e_0 = N\dot{\varphi}_1 + \frac{1}{C_2}\int i_1 dt = N\dot{\varphi}_2 + \frac{1}{C_2}\int i_2 dt \tag{3}$$

where i_s = source current, e_0 = branch voltage, N = number of turns on each core, and C_2 = value of the series capacitor of each of the two branches.

In the calculations which follow, only the fundamental components of fluxes and currents are considered. This, too, is an approximation, of course. To justify the phase of the fundamental component of flux, in the vector diagram of Fig. 5, the following considerations may be helpful: The branch parameters are assumed to be such that the branch with the smaller current is unsaturated and hence highly inductive, and the branch with the larger current, though highly saturated, still remains predominantly inductive. That the phase relations of the various vectors derived from these assumptions are approximately correct is demonstrated by reference to the oscillograms of an actual circuit; see Fig. 3, for example. The two core fluxes are thus found to lead the source voltage by 90 degrees, as shown in Fig. 5, and hence

$$\varphi_1 = A \cos \omega t \qquad \varphi_2 = B \cos \omega t \tag{4}$$

where A and B are the amplitudes of the fundamental components of core fluxes and ω is the angular frequency of the a-c source.

These latter expressions are substituted into equations 2 and 3, with the aid of the trigonometric identity given in Appendix I, equation 70 resulting in

$$I_s = a_1(A+B) + \left(\frac{3}{4}\right)a_3(A^3+B^3) \tag{5}$$

$$0 = (A-B)\left(\frac{a_1}{C_2} - \omega^2 N\right) + \left(\frac{3a_3}{4C_2}\right)(A^3-B^3) \tag{6}$$

For any one given value of I_s there is therefore at least one value each for the amplitudes of the two core fluxes A and B, resulting from the simultaneous solution of the two cubic equations 5 and 6. Moreover, because of the complete symmetry of these equations with regard to A and B, there exists in addition to each set of values

$A = K_1 \qquad B = K_2$

another set of values

$A = K_2 \qquad B = K_1$

The basic condition of bistability is hereby established, if it may be assumed without proof that all other states are unstable. (Whether the limiting conditions here derived are necessary and sufficient to ensure the instability of all other operating points has not been determined thus far. In fact, subsequent analysis shows the existence of other possible pairs of stable operating points under certain conditions. They can be removed by proper design, however, and are here ignored.) Of course, it may happen that the expressions for A and B in terms of the system parameters do not result in real numbers, in a given case, and it is the purpose of this section to establish the limiting conditions, for which two of the solutions to equation 5 and 6 are not real.

In addition to these two sets of core fluxes, there evidently exists a further solution: $A = B$. Its value is determined by solving the following cubic equation for A, which results from equation 5.

$$I_s = 2a_1A + \left(\frac{3a_3}{2}\right)A^3 \tag{7}$$

Inasmuch as the condition $A = B$ does not represent bistability, it may be considered trivial. Physically, it represents the labile equilibrium condition of the bi-

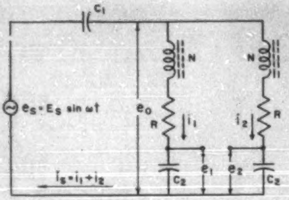


Fig. 4. Schematic diagram of the low-loss ferroresonant trigger pair

stable system, and is therefore of philosophic interest only.

If the case $A=B$ is excluded, in equation 6, the following relation remains

$$0=\left(\frac{a_1}{C_2}-\omega^2 N\right)+\left(\frac{3a_3}{4C_2}\right)(A^2+AB+B^2) \quad (8)$$

It now becomes convenient to introduce some new definitions of terms. Thus

$$L_0=N\frac{d\varphi}{di}\bigg]_{i=0,\ \phi=0}$$

In recognition of equation 1, $i=0$, $\varphi=0$, this expression for the unsaturated inductance of the inductor becomes

$$L_0=\frac{N}{a_1} \quad (9)$$

Further, L_0 and C_2 define a frequency

$$\omega_0^2 \triangleq \frac{1}{L_0C_2}=\frac{a_1}{NC_2} \qquad \frac{NC_2}{a_1}=\frac{1}{\omega_0^2} \quad (10)$$

The factor $3a_3/4a_1$ will recur often and for many manipulations it is conveniently expressed as

$$m \triangleq \frac{3a_3}{4a_1} \quad (11)$$

With the aid of equations 10 and 11, equations 5 and 6 will respectively read as follows if the trivial case $A=B$ is excluded

$$\psi_1 \equiv 0=-\frac{I_s}{a_1}+(A+B)+m(A^2+B^2) \quad (12)$$

$$\psi_2 \equiv 0=-S+m(A^2+AB+B^2) \quad (13)$$

where S is given by

$$S \triangleq \left(\frac{\omega^2}{\omega_0^2-1}\right)=\omega C_2\left(\omega L_0-\frac{1}{\omega C_2}\right) \quad (14)$$

Equations 12 and 13 now need to be solved simultaneously. However, methods such as the broadside attack of eliminating one of the unknowns will result in a sixth degree equation in the other unknown, and only a graphical solution is likely to succeed, with its inevitable loss of generality.

First, equation 13 will be examined more closely. Clearly, it relates all possible values of A and B, and only the introduction of equation 12 provides the additional conditions which pair possible values of A and B. The curve in the A-B plane, given by equation 13 is an ellipse, and after rotation of the axes by 45 degrees, the transformed equation becomes

$$\psi_2 \equiv 0=3\alpha^2+\beta^2-\frac{2S}{m} \quad (15)$$

Thus, the ellipse is seen to have the half-axes

$$a_\alpha=\sqrt{\frac{2S}{3m}} \qquad a_\beta=\sqrt{\frac{2S}{m}} \quad (16)$$

From equation 16 it is immediately evident that a necessary condition for realizability of bistability is

$$S>0 \text{ and hence } \omega>\omega_0 \quad (17)$$

Unfortunately, equation 12 does not suggest a quantitative graphical interpretation. However, it is the coexistence of the ellipse, ψ_2 of equation 15, and the curve ψ_1 in the A-B plane and represented by equation 12, which defines the stable operating points. This is illustrated in Fig. 6, which shows both curves.

From Fig. 6, several important conclusions may be drawn:

1. If the parameters of ψ_1 and ψ_2 are suitably chosen, there may exist four, rather than two, pairs of values of A and B. The limit of this condition occurs if the two curves have a common point at $A=B=\sqrt{S/3m}$ and points P_1 and P_2 coincide. Thus, for bistability, a condition more stringent than equation 17 will need to be met.

2. If large fluxes, and hence currents, are desired in the two branch circuits, parameter m in equation 16 should be made small. This implies, from equation 11, that only a small degree of nonlinearity is desired, in the flux-current relationship of the inductor. Conversely, if the system is to operate with small currents, a highly nonlinear core characteristic should be selected.

3. To determine the limiting values of circuit parameters for which bistability can be realized, the two curves must, in the limit, osculate at another point, besides the point $A=B$.

Next, the limiting parameter relations suggested by statements 1 and 3 ,will be determined.

If ψ_1 and ψ_2 are to osculate, the following expressions must hold

$$\frac{\partial\psi_1}{\partial A}=\frac{\partial\psi_2}{\partial A} \qquad \frac{\partial\psi_1}{\partial B}=\frac{\partial\psi_2}{\partial B} \quad (18)$$

Application of equation 18 to equations 12 and 13 results in, respectively

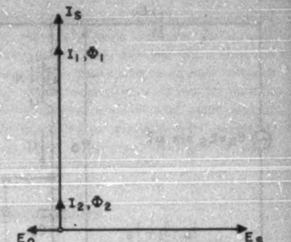


Fig. 5. Vector diagram for the loss-free ferroresonant trigger pair

$$1+3mA^2=m(2A+B) \quad (19)$$

$$1+3mB^2=m(A+2B) \quad (20)$$

Division of these two equations by each other and subsequent simplification result in a curve which relates the A-B coordinates at all points at which a common tangent can be formed

$$3m(A^2-B^2)+(6mAB-1)(A-B)=0 \quad (21)$$

Evidently the straight line $A=B$ also defines such a point, as was anticipated in the foregoing statement 1. Intersection with ψ_2 yields

$$A=B=\pm\sqrt{\frac{S}{3m}} \quad (22)$$

If the constant I_s in ψ_1 is now adjusted so as to place the point of tangency on ψ_1, there follows from equations 12 and 22

$$\boxed{|I_s|<2a_1\left(\frac{S}{3+1}\right)\sqrt{\frac{S}{3m}}} \quad (23)$$

for no more than two stable operating points. Equation 23 thus defines the limit value of I_s. If I_s exceeds this value, the point $A=B$ on ψ_1 moves further away from the origin, and there may exist four stable operating points. If it falls below this value, there will be only two stable operating points.

If the relation of the parameters is not proper, ψ_1 may not exist in the form shown in Fig. 6 and only the trivial condition $A=B$ may be realizable. This aspect will be examined next.

If the condition $A=B$ is excluded, in equation 21 the following relation results

$$\psi_3 \equiv 0=A^2+3AB+B^2-\frac{1}{3m} \quad (24)$$

This expression represents the hyperbola ψ_3 in the α-β plane in Fig. 6. Rotating the axes again by 45 degrees results in

$$\psi_3 \equiv 0 = 5\alpha^2 - \beta^2 - \frac{2}{3m} \qquad (25)$$

The half axes are given by

$$b_\alpha = \sqrt{\frac{2}{15m}} \qquad b_\beta = \sqrt{\frac{2}{3m}} \qquad (26)$$

The ellipse ψ_2 and the hyperbola ψ_3 will intersect if their respective half-axes obey the relation

$$b_\alpha < a_\alpha \text{ or } \sqrt{\frac{2}{15m}} < \sqrt{\frac{2S}{3m}} \qquad (27)$$

Thus, if such an intersection is to exist, thereby permitting a point of osculation for ψ_1 and ψ_2, the following limitation must be imposed

$$\boxed{S > 1/5 \text{ or } \omega^2 > 1.2\omega_0^2} \qquad (28)$$

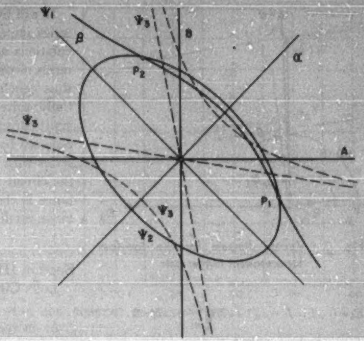


Fig. 6. Graphical representation of the three operating characteristics ψ_1, ψ_2, and ψ_3, which define the response of the ferroresonant trigger pair

Section II. Responses of Loss-Free Circuit

The preceding section dealt with the limiting relations among the circuit parameters, for which a bistable system results. In the calculations which follow, the actual values of the branch currents, output voltage, and supply voltage will be determined for the loss-free system. Again, only the first two terms of the current-flux power series will be considered. It is assumed that the circuit parameters have been so adjusted as to permit bistability.

It was demonstrated in the preceding section that the intersections of ψ_1 and ψ_2 in Fig. 6 represent the stable operating points. For their quantitative evaluation, equations 12 and 13 need to be solved simultaneously. Clearly, a general solution of these two equations is not readily available. However, under certain justifiable assumptions, successive approximations are possible. The two equations are

$$\psi_1 \equiv 0 = -\frac{I_s}{a_1} + (A+B) + m(A^3+B^3) \qquad (29)$$

$$\psi_2 \equiv 0 = -S + m(A^2+AB+B^2) \qquad (30)$$

If the system is to be of practical use, it may be stipulated that one of the stable states shall result in a negligible branch current, the other being large. Expressed otherwise, it is required that the "switching ratio" Δ be as large as possible, the latter being defined as the ratio, of the larger to the smaller current, for either of the two stable states, P_1 or P_2, in Fig. 6. Reference to the oscillograms in section III shows that adequate experimental justification is available for the following statement

$$\begin{array}{lll} B \approx 0 & A \gg 0 & \text{at } P_1 \\ B \gg 0 & A \approx 0 & \text{at } P_2 \end{array} \qquad (31)$$

It will be understood, of course, that such a desirable performance will impose certain further restrictions on the source voltage and current, which will subsequently be determined. In accordance with these statements, flux B in equation 30 is set equal zero, resulting in

$$A = \pm\sqrt{\frac{S}{m}} \qquad B = 0 \qquad (32)$$

If a better approximation is desired, the following procedure may be followed: First, equation 40 is multiplied again, by $(A\text{-}B)$; then equation 29 is equated to the result. After simplification, the following relation is found

$$B = \frac{\left[\left(\frac{I_s}{a_1}\right) + A(S-1) - 2mA^3\right]}{(S+1)} \qquad (33)$$

This value of B could now be introduced into equation 30, and a new, improved value of A be found. This new value, upon substituting into equation 33, will yield a corrected value for B. Thus, by successive approximation, good accuracy can be achieved, provided the solution converges. However, in the light of the rather sweeping assumptions delineated at the beginning of the paper, it would appear that such improvements do not justify the time required to obtain them.

A check on a previous result can now be obtained: If $A = B$ in equation 30, there results, by solving equation 30 for A

$$A = \pm\sqrt{\frac{S}{3m}} \qquad (34)$$

If this value is introduced into equation 29, there follows

$$\pm I_s = 2a_1 \; 1 + \frac{S}{3}\sqrt{\frac{S}{3m}} \qquad (35)$$

Comparison with equation 23 shows this result to be identical with that obtained previously, by a different approach. Thus, to ensure bistability, the source current must not exceed the value given in equation 35. This permits the establishment of a relation between source voltage and branch voltage

$$E_0 + E_s = I_s\left(\frac{1}{\omega C_1}\right) \qquad (36)$$

Hence, for bistability

$$E_s < \left(\frac{2a_1}{\omega C}\right) 1 + \frac{S}{3}\sqrt{\frac{S}{3m}} - E_0 \qquad (37)$$

The desirability of the largest possible switching ratio has been decided upon; it now remains to determine the conditions which permit this ratio to become infinite. Thus, if $B = 0$, the following must hold

$$E_0 = 0 \quad I_1 = I_s \quad I_2 = 0 \quad E_s = \frac{I_s}{\omega C_1} \quad E_1 = \frac{I_s}{\omega C_2} \quad E_2 = 0 \qquad (38)$$

Of interest is the observation that in the vicinity of optimum operation, the smaller of the two branch currents may be made to reverse phase, as I_s varies slightly. Thus, points P_1 and P_2 of Fig. 6 may both fall into the first quadrant of the A-B plane, or else P_2 and P_1 are in the second and fourth quadrants respectively.

Since current and flux are related by

$$i_1 = a_1\varphi_1 + a_3\varphi_1^3 \qquad (39)$$

there follows, for the case $\varphi_1 = A\cos\omega t$, $\varphi_2 = 0$, that the nonzero branch current is as follows

$$I_1 = a_1A + \frac{3a_3A^3}{4} \qquad (40)$$

Then, in recognition of equations 34 and 40, the following relations result by substitution into equations 35 to 38, i.e., for the optimum case.

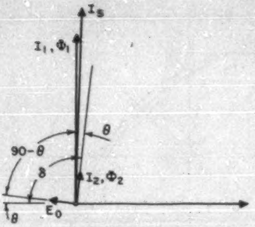


Fig. 7. Vector diagram for the low-loss ferroresonant trigger pair

$$E_0=0 \quad I_1=I_s=a_1(1+S)\sqrt{\frac{S}{m}} \qquad I_2=0$$

$$E_1=\left[\frac{a_1(1+S)}{\omega C_2}\right]\sqrt{\frac{S}{m}}=\left(\frac{a_1\omega}{\omega_0{}^2C_2}\right)\sqrt{\frac{S}{m}}$$

$$=\omega N\sqrt{\frac{S}{m}} \qquad E_2=0 \quad (41)$$

$$E_s=\left[\frac{a_1(1+S)}{\omega C_1}\right]\sqrt{\frac{S}{m}}=\left(\frac{a_1\omega}{\omega_0{}^2C_1}\right)\sqrt{\frac{S}{m}}$$

$$=\omega N\left(\frac{C_2}{C_1}\right)\sqrt{\frac{S}{m}}$$

It can thus be seen that supply voltage and output voltage in the optimum case are related by the equation

$$E_1=E_s\left(\frac{C_1}{C_2}\right) \qquad (42)$$

It is evidently advantageous to make the capacitance ratio C_1/C_2 as large as possible, still remaining, of course, in accordance with equation 36 to 38.

It is quite likely, that a design in accordance with equation 41 will require rather large source voltages, and a compromise between optimum and realistic design may be necessary. In this case the branch voltage will not be zero, and the lower of the branch currents is given to a sufficient approximation by

$$I_2=\frac{E_0}{\left(\omega L_0-\frac{1}{\omega C_2}\right)}=E_0\left(\frac{\omega C_2}{S}\right) \qquad (43)$$

Since the switching ratio can be defined as

$$\Delta \triangleq \frac{I_1}{I_2}=\frac{(I_s-I_2)}{I_2}=\left(\frac{I_s}{I_2}\right)-1 \qquad (44)$$

and since, approximately

$$I_s=\frac{(E_s+E_0)}{\omega C_1} \qquad I_2=E_0\left(\frac{\omega C_2}{S}\right) \qquad (45)$$

there follows, by substituting equation 45 into equation 44

$$\boxed{\Delta=\left(\frac{C_1}{C_2}\right)S\left(\frac{1+E_s}{E_0}\right)-1} \qquad (46)$$

It will be noted that the nature of the series impedance does not enter the calculations as long as all responses are in terms of the source current I_s. One may thus conclude that the circuit is essentially suited to be driven from a constant-current source, but the latter can be replaced by a constant-voltage supply in series with any arbitrary impedance. In the circuit under consideration, a capacitance is chosen, but resistance or inductance would also serve the purpose.

Section III. Responses of Low-Loss Circuit

As in section I, the current-flux relations are assumed to be simply

$$i_1=a_1\varphi_1+a_3\varphi_1{}^3 \qquad i_2=a_1\varphi_2+a_3\varphi_2{}^3 \qquad (47)$$

Corresponding to equations 2 and 3, two relations can be established for the circuit of Fig. 4

$$i_s=i_1+i_2 \qquad (48)$$

$$e_0=N\dot{\varphi}_1+Ri_1+\left(\frac{1}{C_2}\right)\int i_1dt$$

$$=N\dot{\varphi}_2+Ri_2+\left(\frac{1}{C_2}\right)\int i_2dt \quad (49)$$

After differentiation and application of equation 47, equation 49 reads

$$N(\ddot{\varphi}_1-\ddot{\varphi}_2)+Ra_1(\dot{\varphi}_1-\dot{\varphi}_2)+$$
$$3Ra_3(\varphi_1{}^2\dot{\varphi}_1-\varphi_2{}^2\dot{\varphi}_2)+$$
$$(\varphi_1-\varphi_2)\frac{a_1}{C_2}+\left(\frac{a_3}{C_2}\right)(\varphi_1{}^3-\varphi_2{}^3) \quad (50)$$

The next consideration concerns the assumption of a phase for the fundamental components of currents of the two branches, and hence for the two core fluxes. However, once again the calculations will become rather involved, unless certain simplifying assumptions are made. Thus, one may require that the series capacitor be so chosen that

$$\frac{I_s}{\omega C_1}\gg E_0 \qquad (51)$$

This condition is normally met, since E_0 is nearly always negligibly small; I_s is then determined primarily by the reactance of C_1, and leads the source voltage by approximately 90 degrees.

Furthermore, in a practical design the switching ratio should be at least five, so that the larger of the two branch currents is approximately in phase with the source current. As outlined in section I, both branches may still be assumed to be predominantly inductive. This is clearly shown in the oscillograms of Fig. 3, where the fundamental components of the branch currents may be said to lead the source voltage by about 90 degrees and are approximately in phase with each other. Thus, a vector diagram for the low-loss case would appear as shown in Fig. 7.

Fig. 3 shows two complete sets of oscillograms, for an actual circuit. Fig. 3 (A) shows the response for $E_s=100$ volts rms and Fig. 3(B) for $E_s=250$ volts rms, for comparison.

Based upon the preceding evidence, the following assumptions of flux wave equations are reasonably justified, for one of the two stable states

$$\varphi_1=A\cos\omega t \qquad \varphi_2=B\cos(\omega t-\theta) \qquad (52)$$

The phase angle between these two fluxes can be determined as shown in Appendix II, and is given by the equation

$$\theta=\left[\tan^{-1}\frac{(S-mB^2)}{R\omega C_2(1+mB^2)}\right]-90° \qquad (53)$$

It is interesting to note that θ will decrease as B is increased. Thus, a larger source voltage will decrease the phase angle between the two fluxes. This observation is also illustrated by the oscillograms of Fig. 3(A) and (B), if the fluxes may be assumed to be in phase with their respective magnetomotive forces.

The assumption is now made that the circuit should have reasonably low losses, so that θ is a very small angle but not zero. Moreover, the quantity mB^2 in equation 53 shall be small compared to S, and compared to unity. Justifications for these assumptions are found in Appendix II, where their introduction is shown to lead to the following value, for the critical resistance of bistability.

$$\boxed{\begin{array}{l} R<\sqrt{\left(\omega L_0-\frac{1}{\omega C_2}\right)\left(\frac{1}{\omega C_2}\right)} \\ \text{or}\sqrt{\frac{S}{\omega C_2R>1}} \end{array}} \qquad (54)$$

The qualitative implication of equation 54 is important; the closer the operating frequency approaches the natural frequency of each branch in the unsaturated state, the lower the resistance must be kept, to permit bistability.

A rough approximation of the larger of the two core fluxes is now obtained from equation 87 by assuming B^2 and AB to be negligible, compared to A^2, thus

$$A\approx\pm\sqrt{\frac{(S^2+R^2\omega^2C_2{}^2)}{m(S-R^2\omega^2C_2{}^2)}} \qquad (55)$$

The source current, and hence the larger of the two branch currents, is given by a calculation similar to equation 36 to equation 41, resulting in

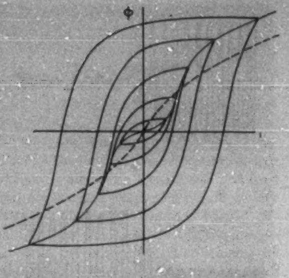


Fig. 8. Nest of hysteresis loops. The loop tips are connected by the magnetization line, of which the broken curve is a useful approximation

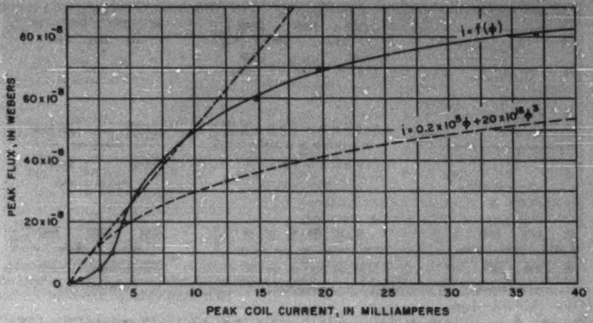


Fig. 9. Experimentally obtained magnetization line I=f(φ), and its 2-term approximation. The two curves have similar slopes, in certain ranges

$$I_s \approx I_1 = a_1(S+1)\sqrt{\frac{(S^2+R^2\omega^2C_2^2)}{m(S-R^2\omega^2C_2^2)}} \qquad (56)$$

The output voltage is found from

$$E_2 = \frac{I_s}{\omega C_2} \qquad (57)$$

where I_s is given by equation 56.

In this evaluation of the source current, it was found desirable to assume both the branch voltage, and the smaller branch current to be zero. A better approximation of the branch voltage, and subsequently of the smaller branch current, can be obtained as follows: from equation 49, after substitution of equation 47 and differentiation, the following relation results for the fundamental components only

$$E_0^2 = \left(\frac{a_1^2}{\omega^2C_2^2}\right)[(-SA+mA^3)^2+ R^2\omega^2C_2^2(A+mA^3)^2] \qquad (58)$$

The approximate value for A given in equation 55 can now be substituted into equation 58, resulting in

$$E_0 = \frac{a_1R(S+1)(S^2+R^2\omega^2C_2^2)}{(S-\omega^2R^2C_2^2)\sqrt{m(S-R^2\omega^2C_2^2)}} \qquad (59)$$

Since the core carrying flux φ_2 is essentially unsaturated, its branch impedance is quite accurately given as

$$Z_2 = \sqrt{R^2+\left(\omega L_0-\frac{1}{\omega C_2}\right)^2} \equiv \left(\frac{1}{\omega C_2}\right)\sqrt{R^2\omega^2C_2^2+S^2} \qquad (60)$$

Thus, the smaller branch current is found from equation 51 and equation 60 as

$$I_2 = \frac{E_0}{Z_2} = \left[\frac{a_1R\omega C_2(S+1)}{(S-R^2\omega^2C_2^2)}\right] \sqrt{\frac{(S^2+R^2\omega^2C_2^2)}{m(S-R^2\omega^2C_2^2)}} \qquad (61)$$

For small resistance, this reduces to

$$I_2 \approx \frac{a_1R\omega C_2(S+1)}{\sqrt{mS}} \qquad (62)$$

The switching ratio is now readily determined, from equations 56 and 61

$$\Delta = \frac{(S-R^2\omega^2C_2^2)}{\omega RC_2} \qquad (63)$$

If the resistance is sensibly small, the second term in the numerator can be neglected with respect to the first term, resulting in

$$\Delta \approx \frac{S}{\omega RC_2} \equiv \sqrt{S}\left(\frac{\sqrt{S}}{\omega RC_2}\right) \qquad (64)$$

Equation 64 shows why a small resistance is most desirable, where large switching ratios are needed.

It is also interesting to note that the switching ratio as here defined, is not a function of the degree of nonlinearity. However, as can be seen from equation 56 and equation 62, both branch currents increase, as m is decreased. Thus, a slightly nonlinear system will carry the larger branch currents, and hence develop higher output voltages, whereas the highly nonlinear system will have small currents, and lower voltages, with the same switching ratio for both systems, as long as the systems are identical with regard to all small-signal parameters.

Section IV. Experimental Results

It is the purpose of this section to state such experimental evidence as has been accumulated pertaining to the ferroresonant flip-flop, regardless of the support or the contradiction which this evidence may offer of the theoretical developments derived in the preceding sections. Also, since the calculations are concerned only with the fundamental components of voltage and currents, a comparison with test data becomes in several instances, a matter of interpretation. Clearly, various further tests (as well as improved experimental techniques, notably the use of a narrow-band filter to select for measurement the fundamental components, and to reject all harmonics) would have been helpful. Alternately, a Fourier analysis of all the complex wave shapes could have been carried out. If, however, this omission remains the major flaw in the developments outlined in the analysis, the author will cheerfully yield this point to the critic.

The following describes the nonlinear inductors used in the experimental setup:

Core material by Syracuse General Electric Company, no. 2.98 and 2.99.

Shape of core: Toroid with rectangular cross section.

Approximate size: Inside diameter: 0.275 inch; outside diameter: 0.39 inch; thickness: 0.07 inch.

Power winding: 50 random-wound turns of four strands of Formex-insulated wire, size 26. The resultant four coils are connected in additive series, forming a 200-turn coil, self-resonant at about 500 kc.

Trigger coil: 100 random-wound turns of Formex-insulated wire, size 36, on a 9/16-inch circular coil form, surrounding the toroid. Axis of trigger coil and axis of toroid are at right angles to each other. The trigger coils of the two branches are series-connected and supplied with trigger pulses.

The first step in the performance prediction of the ferroresonant trigger pair is, of course, the determination of the coefficients of the current-flux relation of equation 47. Several methods are outlined in Appendix III, and it is sug-

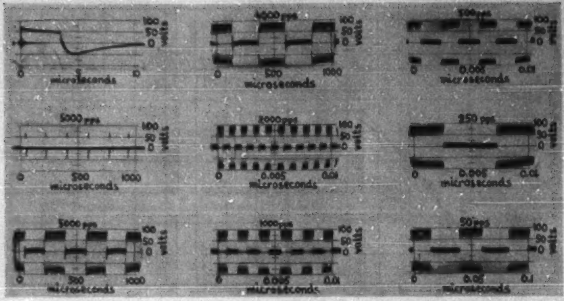


Fig. 10. Response of the ferroresonant trigger pair to 3.5-microsecond 65-volt trigger pulses, at trigger rates ranging from 5,000 to 50 pulses per second. $E_s = 50$ volts; $f = 233$ kc

gested there that a dynamic measurement would yield more reliable data. Accordingly, one of the two identical cores with 200 turns each was connected in series with a 10-ohm resistor and an adjustable voltage source of 233 kc, and of sinusoidal wave shape. The peak value of flux in the core for any given value of source voltage is given, to a sufficient approximation, by the relation

$$\varphi_{\text{peak}} = E_{\text{rms}} \frac{10^8}{4.44fN} = 0.483E_{\text{rms}} \qquad (65)$$

The peak value of exciting current corresponding to any one value of source voltage is proportional to the voltage drop across the 10-ohm resistor in series with the coil, and a plot of peak flux versus peak current thus obtained yields a "magnetization line" which connects the tips of a nest of hysteresis loops, as shown in Fig. 8. The result of this procedure is shown in Fig. 9.

It is readily seen from Fig. 9 that a satisfactory 2-term approximation is not possible, if the actual current-flux relation is to be simulated over a sufficiently wide range.

The reader can convince himself readily that very widely different choices of the two coefficients of equation 65 are possible all of which appear to have merits over certain ranges. Indeed, an improvement is sometimes possible by allowing the first coefficient to become negative, which implies a negative value of inductance for very small currents, and a physical justification of such a term becomes rather tenuous. Once more, then, it appears wise to take stock of the physics of the device before proceeding with the mathematics.

The significant circuit parameter is the rate at which flux varies with current; this quantity is, of course, related to the differential inductance by the number of turns, as seen from equation 66, and it is this same differential inductance which governs the manner in which the instantaneous current through the coil varies with the voltage across the coil.

$$e_{L_1} = N\frac{d\varphi}{dt} = N\left(\frac{d\varphi}{di}\right)\left(\frac{di}{dt}\right) = L_1\frac{di}{dt} \qquad (66)$$

Thus, an approximation appears indicated in which the slopes near the origin and over a wide part of the saturation range approximate their respective experimental equivalents. Such an approximation is shown in Fig. 9, and is governed by the relation.

$$i = 0.2\times10^5\varphi + 20\times10^{15}\varphi^3 \qquad (67)$$

It is now obvious why, in the author's opinion, an enhancement of mathematical accuracy through choice of a third term in the current flux relation, becomes meaningless.

The coil resistance is negligible and, from this choice of the coefficients a_1 and a_3, the following data are obtained (for $N = 200$, $f = 233$ kc, and $C_1 = C_2 = 500 \times 10^{-12}F$)

$$\omega_0^2 = \frac{a_1}{NC_2} = 0.2\times10^{12} \qquad \omega = 1.46\times10^6$$

$$\omega^2 = 2.14\times10^{12}$$

$$S = \left(\frac{\omega^2}{\omega_0^2} - 1\right) = 9.7 \qquad m = \frac{3a_3}{4a_1} = 7.5\times10^{11}$$

$$\sqrt{\frac{S}{m}} = 1.16\times10^{-6} \qquad (68)$$

For the optimum case the following results were obtained

$$E_0 = 0 \quad E_s = \left(\frac{\omega NC_2}{C_1}\right)\sqrt{\frac{S}{m}} = 333 \text{ volts peak}$$

$$E_1 = \omega N\sqrt{\frac{S}{m}} = 333 \text{ volts peak} \quad E_2 = 0 \qquad (69)$$

$$I_1 = I_s = a_1(1+S)\sqrt{\frac{S}{3m}} = 0.14\text{-ampere}$$

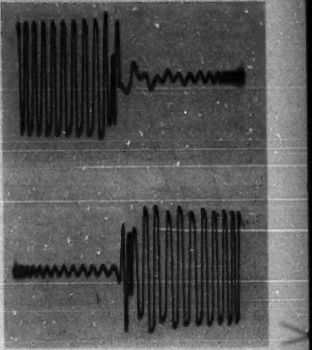


Fig. 11. State changes of the ferroresonant trigger pair. Approximately 5 cycles are needed to change from the high-current state to the low-current state, or vice versa

$$\text{peak} \qquad I_2 = 0 \qquad \Delta = \frac{I_1}{I_2} = \infty$$

The source voltage dictated by optimum considerations is rather large, resulting in large currents and excessive core heating. However, comparable results are obtainable with smaller source voltages. Oscillograms for two such cases were shown in Fig. 8, and approximate numerical results are listed in Table I. (It should be noted once more that a Fourier analysis of the curves of Fig. 3(A) and (B) is needed to obtain the exact peak values of the fundamental components of voltages and currents.)

The switching ratios resulting from the two source voltages listed in Table I are calculated from the ratio of the fundamental components of branch currents or, if preferred, from the ratios of the fundamental components of output voltages. They can also be calculated as follows

Source voltage E_s, volts peak	100	250
Calculated $\Delta = \left(\frac{1+E_0}{E_0}\right)\frac{SC_1}{C_2}$	28.1	44
Measured Δ	20	31

The response of the ferroresonant trigger pair to various trigger pulse rates is shown in Fig. 10. Due to pulse rate limitations on the equipment supplying the trigger signals, the limit on circuit response could not be determined readily. However, as seen from the oscillograms of Fig. 11, approximately five cycles of the carrier frequency are needed for a change of state of the circuit. Thus, for the circuit tested, the ultimate trigger rate may be expected to be of the order: trigger rate = 1/5 (frequency).

It is interesting to note that the use of a trigger circuit of low impedance lowers the source voltage necessary for bistability. This is readily understood, however, from an inspection of equation 41. Clearly, the smaller the quantity S becomes, the lower the necessary source voltage. Since the trigger coil and load coil are coupled magnetically, albeit only slightly, the low-impedance trigger circuit has the effect of reducing the inductance of the power coil, thus raising its unsaturated resonant frequency. Since this results in a reduction of S, the reduction of source voltage observed in taking the data of Figs. 10 and 11 appears satisfactorily explained.

Appendix I

The following three trigonometric identities have been used in the preceding analysis

$$(\cos x)^{2n+1} \equiv (1/2^{2n})[\cos(2n+1)x + (2n+1)\cos(2n-1)x + \binom{2n+1}{2}\times \cos(2n-3)x + \binom{2n+1}{3}\times \cos(2n-5)x + \ldots\binom{2n+1}{k-1}\times \cos(2n-2k+3)x\ldots] \quad (70)$$

$$\sin x \cos^2 x \equiv (1/4)\sin x + (1/4)\sin 3x \quad (71)$$

$$\cos(\omega t-\theta) \equiv \cos\omega t \sin\theta + \sin\omega t \cos\theta \quad (72)$$

$$\sin(\omega t-\theta) \equiv \sin\omega t \cos\theta - \cos\omega t \sin\theta$$

Appendix II

The angle θ can be determined from equation 47, as the phase angle relating the in-phase and quadrature components of φ_2 as follows

$$\dot{e}_0 = \left(N\ddot{\varphi}_2 + \frac{i_2}{C_2}\right) + R\dot{i}_2 \quad (73)$$

If the current-flux relation of equation 47 is introduced, there results:

$$\dot{e}_0 = \left[N\ddot{\varphi}_2 + \frac{(a_1\varphi_2 + a_3\varphi_2^3)}{C_2}\right] + [Ra_1\dot{\varphi}_2 + 3Ra_3\varphi_2^2\dot{\varphi}_2] \quad (74)$$

This equation converts to the following, with the aid of trigonometric identities of equations 70 and 71, Appendix I, provided only the fundamental components are related, as

$$\dot{e}_0 = \left[\left(-N\omega^2 + \frac{a_1}{C_2}\right)B + \frac{3a_3B^3}{4C_2}\right]\cos(\omega t-\theta) - R\omega\left(a_1B + \frac{3a_3B^3}{4}\right)\sin(\omega t-\theta) \quad (75)$$

With the introduction of the terms

$$\omega_0^2 \triangleq \frac{a_1}{NC_2} \qquad m \triangleq \frac{3a_3}{4a_1} \qquad S \triangleq \left(\frac{\omega^2}{\omega_0^2} - 1\right) \quad (76)$$

the phase of $\dot{e}_0$ is thus given by

$$\tan\delta = (S - mB^2)/R\omega C_2(1+mB^2) = \left[\frac{\left(\frac{\omega^2}{\omega_0^2}\right)}{R\omega C_2(1+mB^2)}\right] - \frac{1}{R\omega C_2} \quad (77)$$

The angle can now be found from

$$\theta = \left[\tan^{-1}\frac{(S-mB^2)}{R\omega C_2(1+mB^2)}\right] - 90° \quad (78)$$

Since the two branches are in parallel, the two branch voltages must be identical, and the following equality must hold

$$(-SB+mB^3)\cos(\omega t-\theta) - R\omega C_2(B+mB^3)\sin(\omega t-\theta) = (-SA+mA^3)\cos\omega t - R\omega C_2(A+mA^3)\sin\omega t \quad (79)$$

With the use of the trigonometric identities 72 in Appendix I, this latter equation can be expanded, and two independent relations result from equating the sine and cosine expressions, respectively:

$$(-SB+mB^3)\cos\theta + R\omega C_2(B+mB^3)\sin\theta = (-SA+mA^3) \quad (80)$$

$$(-SB+mB^3)\sin\theta - R\omega C_2(B+mB^3)\cos\theta = -R\omega C_2(A+mA^3) \quad (81)$$

In recognition of the assumption of a low-loss circuit, and the evidence presented in the oscillograms, the angle θ is nearly zero degrees, and hence:

$$\cos\theta \approx 1 \quad (82)$$

Thus, equations 80 and 81 can be simplified:

$$-S(A-B) + m(A^3-B^3) = R\omega C_2B(1+mB^2)\sin\theta \quad (83)$$

$$R\omega C_2(A-B) + R\omega C_2m(A^3-B^3) = -B(-S+mB^2)\sin\theta \quad (84)$$

Dividing these two equations by each other results in:

$$[-S+m(A^2+AB+B^2)]/R\omega C_2[1+m(A^2+AB+B^2)] = -R\omega C_2(1+mB^2)/(mB^2-S) \quad (85)$$

Equation 85 reduces considerably through the assumption that: $mB^2 \ll 1$

This is not unreasonable: Equation 47 can be rewritten, for the case of cosinusoidal flux, and considering fundamental components only:

$$i_2 = Ba_1(1+mB^2) \quad (86)$$

Thus, since i_2 is assumed to be the smaller of the two branch currents, the corresponding core can be considered as linear, the second term can be neglected in comparison to the first, and $mB^2 \ll 1$ is justified. A further assumption is only conditionally permissible: It is desirable that: $mB^2 \ll S$. Here a severe error is introduced if the unsaturated natural frequency of each branch is very nearly equal to the source frequency. However, again in most practical cases, the factor S will probably exceed ten per cent whereas, for the unsaturated condition where B is small, $mB^2 \ll 0.01$, in all likelihood.

Table I. Approximations from Fig. 8

Volts Peak				Amperes Peak			
E_s	E_0	E_1	E_2	I_1	I_2	I_s	Δ
			Case 1				
100	50	80	4	0.10	0.004	0.10	20
			Case 2				
250	70	180	6	0.14	0.0045	0.13	31

After these simplifications, equation 85 is now solved for (A^2+AB+B^2):

$$A^2+AB+B^2 = (S^2+R^2\omega^2C_2^2)/m(S-R^2\omega^2C_2^2) \quad (87)$$

Clearly, this expression reduces to equation 30, for the loss-free circuit:

$$A^2+AB+B^2 = S/m \quad (88)$$

The elliptic locus given by equation 87 ceases to exist if the term on the right side becomes negative. Thus, one may establish the following limitation on the resistance, if bistability is to exist:

$$R < \left(\frac{1}{\omega C_2}\right)\sqrt{\left(\frac{\omega^2}{\omega_0^2} - 1\right)}$$

Rewritten this expression states:

$$R < \sqrt{\left(\omega L_0 - \frac{1}{\omega C_2}\right)\left(\frac{1}{\omega C_2}\right)} \quad \text{or} \quad \frac{\sqrt{S}}{\omega C_2 R} > 1 \quad (89)$$

Appendix III. Some Aids in Determination of Nonlinearity Constants

By supposition, the current-flux relation was given as

$$i = a_1\varphi + a_3\varphi^3 \equiv \varphi(a_1 + a_3\varphi^2) \quad (90)$$

For any given current i the differential inductance is given by $L_i \triangleq \frac{Nd\varphi}{di}$

Thus, it follows from equation 90 that

$$L_i = \frac{N}{a_1} + a_3\varphi^2 \quad (91)$$

Thus, the initial inductance is given by

$$L_0 = \left(N\frac{d\varphi}{di}\right)\bigg]_{i=0} = \frac{N}{a_1} \quad (92)$$

and hence the linearity coefficient becomes

$$a_1 = \frac{N}{L_0} \quad (93)$$

If equaton 93 is substituted into equation 91, and the result is solved for the flux, there follows

$$\varphi = \sqrt{\left(\frac{N}{3a_3}\right)\left[\left(\frac{1}{L_i}\right) - \left(\frac{1}{L_0}\right)\right]} \quad (94)$$

Substitutions of equations 93 and 94 into equation 90 and solving for the nonlinearity coefficient, results in

$$a_3=\left(\frac{N^3}{27i^3}\right)\left[\left(\frac{1}{L_i}\right)-\left(\frac{1}{L_0}\right)\right]\times \left[\left(\frac{1}{L_i}\right)+\left(\frac{2}{L_0}\right)\right]^2 \tag{95}$$

Since the curve of L_i versus i is in general obtained by measurement, the closeness of the fit of the curve given by equation 90 and the coefficients of equations 93 and 95, can be gauged by plotting the current-inductance relation obtained through the use of the coefficients. This information is contained in equation 95 and results in the following

$$i=\left(\frac{N}{3}\right)\left[\left(\frac{1}{L_i}\right)+\left(\frac{2}{L_0}\right)\right]\times \sqrt{\left(\frac{N}{3a_3}\right)\left[\left(\frac{1}{L_i}\right)-\left(\frac{1}{L_0}\right)\right]} \tag{96}$$

A much simpler and more direct method for determining the current-flux coefficients, than the bridge method for determining inductance is based on the availability of calibrated hysteresis loop equipment. Thus, if a "nest" of hysteresis loops is taken, for either of the two identical cores, the loop tips can be connected by the magnetization line. Although the low magnetomotive force portion of this curve often deviates considerably from the relation assumed in equation 90, an approximation as shown in Fig. 9 is not unreasonable. However, an alternate simple method for obtaining the magnetization line is stated at the beginning of section IV.

Once the approximation magnetization line is determined, the desired coefficients are readily found. Thus, as in equation 92

$$\frac{1}{a_1}=\left(\frac{d\varphi}{di}\right)\Big]_{\varphi=0} \tag{97}$$

With a_1 determined, a point near the knee of the magnetization line yields the necessary data to solve for the nonlinearity coefficient a_3. Alternately, two points may be chosen on the magnetization line, resulting in two simultaneous equations

$$\begin{pmatrix}\varphi_a & \varphi_a^3\\ \varphi_b & \varphi_b^3\end{pmatrix}\begin{pmatrix}a_1\\ a_3\end{pmatrix}=\begin{pmatrix}i_a\\ i_b\end{pmatrix} \tag{98}$$

It is obvious that the hysteresis loop method yields more reliable results than that based on the curve of differential inductance versus current; the loops can, and generally should, be taken at the operating frequency of the flip-flop, and thus they very nearly represent actual dynamic operating conditions. In contrast, the differential inductance is essentially based on a d-c hysteresis loop and, as mentioned earlier, the resulting error can be severe. The unwary experimenter should be cautioned, however, that the hysteresis loop equipment should have a bandwidth of no less than 20 times the operating frequency along the magnetomotive force axis with the relative phase shift between the flux axis and the magnetomotive force axis being negligible, over this bandwidth. If these precautions are not observed, the hysteresigram may be considerably distorted. This effect is particularly noticeable in the case of core materials with pronounced nonlinearities.

References

1. Ferroresonant Flip-Flops, **C. Isborn.** *Electronics*, New York, N. Y., Apr. 1952, pp. 121–28.

2. Nonlinear Resonant Flip-Flop Circuit, **C. F. Spitzer,** *et al.* *U. S. Patent no. 2,653,254*, Sept. 22, 1953.

3. Über Neue Resonanzerscheinungen in Wechselstromkreisen, **O. Martienssen.** *Physikalische Zeitschrift*, Leipzig, Germany, vol. 11, 1910, pp. 448–60.

4. **F. Waibel.** *Annalen der Physik*, Leipzig, Germany, vol. 22, 1923, p. 3.

5. Erzwungene Schwingungen bei Veranderlicher Eigenfrequenz (book), **G. Duffing.** F. Vieweg & Son, Brunswick, Germany, 1918.

6. On the Anomalous Behaviour of a Galvanometer, **E. V. Appleton.** *Philosophical Magazine*, London, England, vol. 47, 1924, pp. 609–19.

7. Applied Mathematics for Engineers and Physicists (book), **L. A. Pipes.** McGraw-Hill Book Company, Inc., New York, N. Y., 1947.

8. Nonlinear Vibrations (book), **J. J. Stoker.** Interscience Publishers, New York, N. Y., 1950.

9. Nonlinear Mechanics (book), **N. Minorski.** J. W. Edwards, Ann Arbor, Mich., 1947.

10. Introduction to Nonlinear Mechanics (book), **N. Kryloff, N. Bogoliuboff.** Princeton University Press, Princeton, N. J., 1947.

11. Ferroresonant Circuits for Digital Computers, **C. F. Spitzer.** *Dissertation* for Ph.D. in engineering, Yale University, New Haven, Conn., 1955.

12. The Ferroresonant Trigger Pair: Analysis and Design, **C. F. Spitzer.** *T.I.S. 54-E-292*, General Electric Company, Syracuse, N. Y., Nov. 21, 1954.

13. Analysis and Design of the Ferroelectric Resonant Trigger Pair, **C. E. Gremer.** *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 11, 1955, p. 828.

Transmission of Business Machine Data Over Standard Telegraph Channels

F. B. BRAMHALL
FELLOW AIEE

Computer Requirements

THE transmission of digital information over long distances for the operation of business machines is a problem that will engage increasing attention of communication engineers. High-speed electronic computer-type business machines are being called upon more and more to operate on information fed from distant points. This new form of transmission results in encoded information which is not at all intelligible in the accustomed sense. Such communication is nevertheless telegraphy in every true sense of the word. The record of this new form of record communication may not be a printed page but rather a record more easily reread into the computer machine when needed. The technical problems encountered in the transportation of business-machine information from one point to another are those long familiar to the telegraph engineer.

While it may be expedient that the computer functions of the business machine be carried out at high speed, the operating speed of the communication circuits which bring information from distant associated business offices may be quite different. The reasons for this difference are varied. For instance, the computer may be serving several distant branches on a time-sharing basis. Further, a considerable part of the information fed to the computer for a given task may be unchanged from a recent previous assignment and may, therefore, have already been stored in memory devices. For instance, in payroll work, the hourly rates, the tax rates, and the like are probably unaltered, and only the hours information is brought in over the communication circuit each pay period. Likewise, in an inventory type of operation, prices are remembered by the machine, and only type numbers, quantities, and the like come over the wire. The required telegraphic speed then, instead of being in the region of hundreds of thousands of pulses (bits) per second, may well be such that it can be accommodated by a few hundred cycles of bandwidth or even by one or more conventional telegraph circuits. Fortunately, modern record communication facilities lend themselves readily to

---

Paper 56-654, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 27, 1956; made available for printing May 2, 1956.

F. B. Bramhall is with the Western Union Telegraph Company, New York, N. Y.

almost any signaling speed that ultimately may be proved most advantageous from the standpoint of the machine and the work it is required to perform. Only minor new developments in terminal equipment are required to accommodate these new business methods as they evolve, in any case.

Present Facilities

The 3-kc voice-width band is the vehicle of all present-day, long-haul telegraphy. Operated by time-division technique, with the coded data transmitted seriatim at high speed, this universal vehicle will convey many hundred information bits per second. It may well evolve that this commonly available unit of communication spectrum will come into wide use for feeding information to and delivering results from digital computer-type business machines. Another readily available communication spectrum unit is the half-band. The 3-kc bands of the Western Union Telegraph Company's entire trunk network are divided into two half-bands or subbands, as they are called. Each of these subbands has a useful spectrum range from approximately 350 to 1,625 cycles. This subband with a probable speed capability of perhaps 750 or 1,000 bits per second may prove to be a useful and economical spectrum unit for data transmission.

As used in conventional message traffic and private lease service, each subband carries nine telegraph channels. The signaling rate on these telegraph channels is 45, 57, or 75 bands (bits per second), depending on the nature of the service. Contemplating the widespread use of 100 speed (words per minute), particularly in private wire service of special classes, led to the design of the modern telegraph plant to render substantially errorfree transmission at 75 bits per second. Presently, a considerable demand for data transmission at rates materially lower than 750 bits per second is being met by the use of one or more of these standard trunk telegraph channels. The characteristics and capabilities of these channels are described in the following.

Frequency Modulation

The Western Union Telegraph Company pioneered in the development and application of frequency modulation for telegraphy. Its whole plant, comprising several million channel miles, is operated by this method. As a result, it enjoys a unique freedom from errors caused by changes in transmission equivalent on the vehicle channels and errors caused by extraneous disturbance from all sources.[1,2]

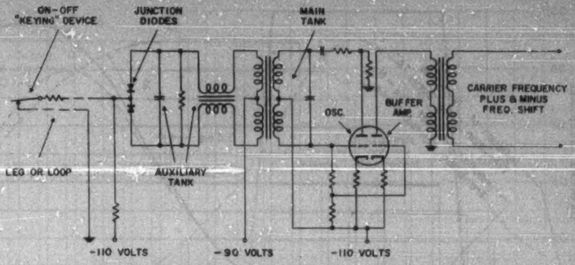


Fig. 1. Electronic transmitter of frequency-modulation telegraph channel

These advantages accrue from the circumstance that positive control is maintained at all times. In transmission, a YES bit is represented by a certain frequency, and a NO bit by another frequency. This contrasts strongly with amplitude modulation, where YES is represented by the presence of a carrier, and NO is only presumed because nothing is transmitted during the intervals of all NO bits. In the amplitude-modulated, ON-OFF method, spurious impulses caused by either man-made or natural interference may be easily interpreted as YES during a NOTHING period. A second great advantage of frequency modulation also follows from the fact that YES and NO are represented by equal-valued signals of different frequencies. Changes in signal strength do not cause shortening or lengthening of the YES bit with respect to the NO. The length of each is determined only by the time interval between the frequency shifts from one to the other and not by the magnitude of the incoming envelope of carrier compared to a local reference level, as in amplitude-modulated signaling. Design objectives and automatic controls notwithstanding, attenuation changes on the vehicle carrier system do occur. Balancing, shielding, and other precautions mitigate but never completely eliminate noise. The transmission system which side-steps these hazards offers great advantage in error freedom.

The Telegraph Channel Transmitter

A brief exposition of the theory and circuitry of the standard telegraph channel terminal shows its extreme simplicity. From a read-out at the sending station, it is required only that d-c potential be applied to a loop circuit which shifts the frequency of the transmitting oscillator. The schematic arrangement of the control circuit of that oscillator is shown by Fig. 1.

The transducer which converts the coded d-c signals into corresponding frequency changes of the carrier is, in fact, only a pair of contact rectifiers.

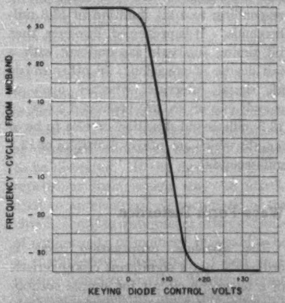


Fig. 2. Response of frequency-modulation oscillator to keying potential

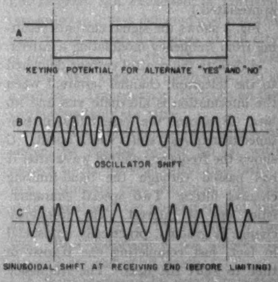
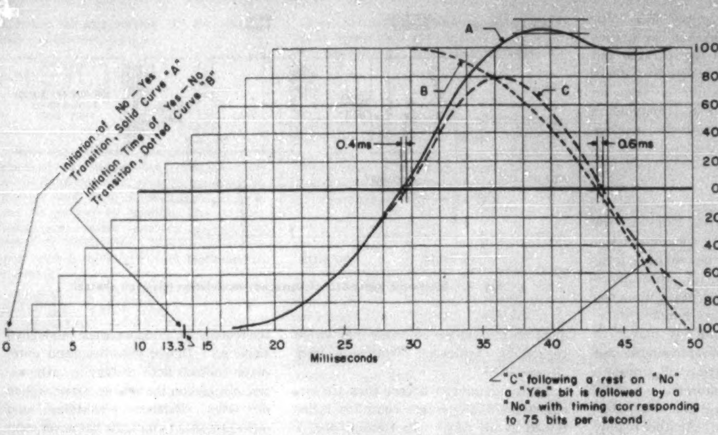

A
KEYING POTENTIAL FOR ALTERNATE "YES" AND "NO"
B
OSCILLATOR SHIFT
C
SINUSOIDAL SHIFT AT RECEIVING END (BEFORE LIMITING)

Fig. 3. The carrier signals in frequency-modulation telegraphy



"C" following a rest on "No" a "Yes" bit is followed by a "No" with timing corresponding to 75 bits per second.

Fig. 4. Time delays and wave forms in telegraph channel at 75 bits per second

When rendered conducting by the opening of the transmitter loop, the frequency of the carrier oscillator is increased. When made non-conducting by the closing of the loop, the oscillator frequency is depressed. The difference between these two frequencies is 70 cycles. The shift from one to the other takes place rather abruptly, with the trigger action specified by Fig. 2, all within the limits of a 20-volt swing in control potential. The frequency shift as observed at the output of the sending-end channel filter, however, is gradual, as dictated by the bandwidth of that filter. (The concept of amplitude time constant in a bandpass filter is a familiar one. It is commonly remembered that the time required to establish a new carrier amplitude at the output of a filter is inversely proportional to half the bandwidth. The fact that a filter has the same reluctance to accommodate a shift in frequency is equally valid, though perhaps not so generally appreciated.)

Fig. 3 shows the signal currents resulting from frequency-modulating a carrier. *A* depicts the control current delivered to the telegraph channel terminal when the information is alternate YES and NO bits. *B* shows the essentially instantaneous frequency shifts which result. *C* shows the frequency-shifted wave after it has passed through the band-limiting channel filters. Two special characteristics are of interest. The frequency change is gradual, practically sinusoidal in fact, just completing one transit in time to start the next. The frequency change is accompanied by a decrease in amplitude corresponding to each upward and downward shift. This latter characteristic is the result of eliminating products of modulation higher than the first-order side bands.

Keying Methods

Actually, a carrier transmitter may be keyed by either of several techniques. A preferred method is the opening and closing of a 2-wire loop, as indicated by Fig. 1. A simpler method, permissible when the keying machine is close to the carrier channel terminal, is the opening and closing to ground of a single wire leg. By rearrangement of the potentials applied to the channel terminal, it is permissible to key by the application of positive battery to a single wire leg. Polar battery too, with positive representing YES and negative representing NO, may be employed.

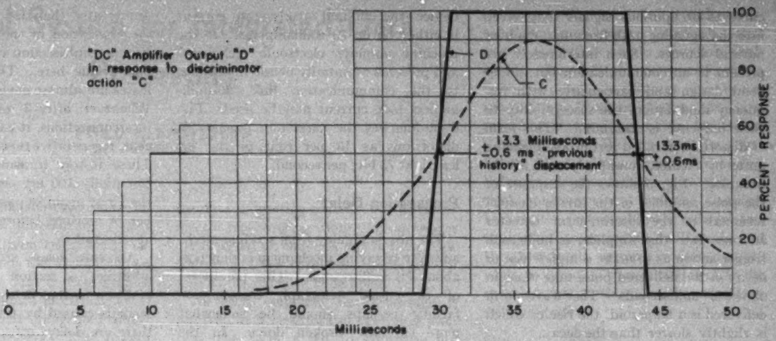
The keying may be at any rate up to and including 75 bits per second, corresponding to a bit duration of 13.3 milliseconds. It is presumed, however, that bit timing intervals are uniform with a fair degree of precision. In other words all signal intervals, whether open or closed, YES or NO, are presumed to be of 13.3-millisecond duration or else to be of lengths which are integer multiples of 13.3 milliseconds.

Wave-Form Distortion

The absolute propagation time is ordinarily not a significant characteristic of a telegraph circuit for simple station-to-station teleprinter working. Only when a communication system depends upon answer-back or error correction does transmission delay become an important matter. Relative shifts of pulse arrival time from the idealized are, however, extremely important in transmission of all forms of encoded information. In nearly all systems of record communication, some selection or gating device functions on a local receiving-end time base. Time displacement of the rise and decay transients of signaling pulses or bits from their idealized position is distortion. Measured in milliseconds, it has a concrete meaning to the terminal equipment designer. For these reasons, considerable attention is devoted to arrival wave forms and the possible conditions which may displace or alter them. Fig. 4 shows time graphs of the response of the channel as measured at the frequency-modulation detector, the discriminator. The response is shown as having a negative sign for the NO or high-frequency condition, and a positive sign for the YES or low-frequency condition. For the immediate purpose, the receiving mechanism will be presumed to be actuated when the detector voltage passes through zero. By this token, it will be noted from graph *A* that the arrival of a YES indicating crossover lags its initiation at the sending station by just under 30 milliseconds. It should be noted also that arrival wave *A* overshoots 100-per-cent response and oscillates mildly.

For the purpose of the following dissertation on wave forms it is convenient to consider a specific pattern of YES and NO bits. The presumed transmitted signal combination for curve *A* is a number of NO bits followed by a number of YES bits. If the situation is reversed and

Fig. 5. Triggered response of channel amplifier output



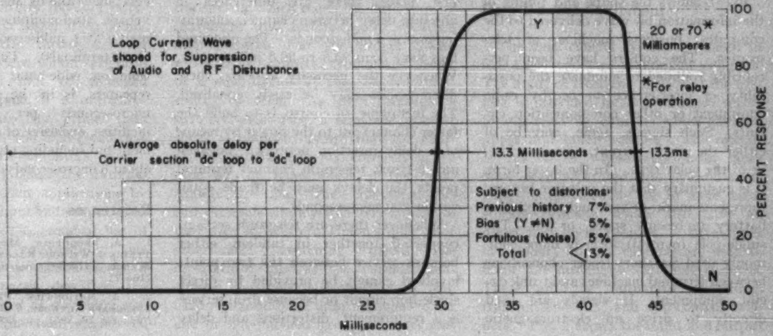


Fig. 6. Shaped signal delivered to data machine

the signals under study are a number of YES bits followed by a number of NO bits, then the transient from one to the other has the shape of curve *B*. The transition from the one to the other is presumed to have been initiated 1-bit interval later than the transition shown by *A*. Ideal transmission would, of course, require that the *A* and *B* arrivals, the points of zero crossing, be precisely 1-bit length apart. This would be only a necessary but not a sufficient criterion of perfect receiving-end signal reproduction. In perfect transmission, the arrival of any bit transient, no matter what information preceded YES or NO, must coincide with perfect bit interval timing. Any departure from ideal timing of a transient caused by the combination of YES and NO pulses may very appropriately be called previous-history distortion.

Graph *C* shows the arrival wave resulting from the transmission of a single YES bit following a number of NO bits. It is to be noted that the zero crossovers marking the beginning and ending of the pulse do not fall precisely where they would had the information content of the character been different. It is not at once apparent which timing is ideal. Speculation quickly reveals that neither beginning nor ending is the same as it would have been had the pulse train been alternate single YES and NO bits. Still, it is not obvious which timing should be called the standard. By accepted usage, the locations of the crossovers of a steady succession of alternate positive and negative bits are taken as the reference standard. Pulses shortened, lengthened, or displaced from that standard are said to be distorted. If the distortion is of a repetitious form which may be attributed to an electrical characteristic of the circuit, sluggish response, or other exhibits of nonlinearity or memory, it is said to be a characteristic distortion. Good engineering practice in the Western Union Telegraph Company holds this distortion within ± 1.0 millisecond per telegraph section.

A telegraph section is defined as that part of a circuit included between two carrier channel terminals. The section length has little, if any, relation to circuit miles. Even many transcontinental circuits are but a single section. The audio-frequency, frequency-shifted carrier generated in New York City for a Los Angeles, San Francisco, or Portland transmission is first detected and restored back to bit-length d-c pulses at the city of destination. On the other hand, the 90-mile New York-to-Philadelphia hop is a section. Transmission efficiency on the two is essentially the same.

Distortion which consistently lengthens either the YES bit or the NO bit with respect to the other is called bias distortion. Marking bias lengthens the YES at the expense of the NO, and spacing bias lengthens the NO at the expense of the YES. Bias distortion is not allowed to exceed 5 per cent of bit length.

A third source of bit-length variation is the distortion caused by noise, crosstalk, and/or whatever form of extraneous currents or spurious signals may creep into the signaling path from man-made or natural sources. Such interferences are present in all communication but, under good design and particularly with frequency modulation, the susceptibility to their influence is minimized. This form of distortion is called fortuitous, meaning quite by chance, completely and wholly irregular. Fig. 5 shows the response of the pulse amplifier of the carrier channel terminal to the discriminator detector action. Into that amplifier is built such trigger action as to make complete rise or decay of the delivered pulse take place in about 2 milliseconds. The wave form delivered is a trapezoid, the rise of which is slightly slower than the decay.

Fig. 6 shows the shape and timing of the information bit wave delivered to the computer, business machine, or teleprinter. The corners have been perceptibly rounded to eliminate the possibility of interference to nearby radio equipment or other communication circuits. Such signals, again, may be of either the single current, open or close, or of the polar type. In the latter form, it is customary that the negative battery represents mark, or YES, and the positive battery represents space, or NO. The single-pole (neutral) form is more commonly used for short feeder connections between terminal machines and trunk circuit equipment. If signals are used directly to drive an electromagnetic device, the nominal steady-state current is either 60 or 70 milliamperes. If the machine employs electronic techniques and presents essentially ohmic impedance to the communication link, 20-milliampere loop current may be used. Fig. 6 summarizes the maximum permissible distortions as 13 per cent of the bit length at 75 bits per second.

Propagation Delay

The absolute delay from sending mechanism to receiving mechanism is shown as about 30 milliseconds. This parameter of the usual telegraph transmission facility perhaps should be somewhat more carefully broken down. In the first place, there are differences in absolute delay between channels amounting to over 4 milliseconds. The measured lags vary from 25.3 to 29.6 milliseconds. Whenever the necessity appears, these discrepancies may be easily equalized. The technique, of course, is to build the faster channel out to the slower by means of a delay insertion network. Between any 2-trunk telegraph channel terminal points the delays may be made equal within 200 microseconds.

Whenever there are telegraph sections connected together in tandem, either because service between the two points involved cannot be provided by direct single-hop circuit or because drop service is a requirement, distortions and delay are greater than for the single section. Generalization in the amount of distortion on multisection circuits would be of little value here. The separate section losses are almost never directly additive. Whenever, after 3- or 4-tandem section interconnections, it exceeds about 30 per cent, regenerative repeaters are employed. These restore transmission efficiency to essentially 100 per cent, and the signals then may again be passed through a number of sections before becoming materially degraded.

Absolute delay, of course, is directly additive to section after section. In addition to that delay of about 30 milliseconds caused by telegraph equipment, there are delays caused by vehicle band equipment and by line propagation. The vehicle band equipment may be approximated at 1 millisecond per band section (two terminals). The line propagation time on wide-band systems, including repeaters, is in the neighborhood of 8 microseconds per mile. Microwave facilities, exclusive of channelizing equipment but including repeaters, account for about 6 microseconds delay per mile.

References

1. A FREQUENCY MODULATION TELEGRAPH TERMINAL WITHOUT RELAYS, **F. H. Cusack, A. E. Michon.** *AIEE Transactions*, vol. 66, 1947, pp. 1165–71.

2. A NATIONWIDE FREQUENCY-MODULATION TELEGRAPH NETWORK, **F. B. Bramhall, L. A. Smith.** *Ibid.*, vol. 70, 1951, pp. 373–78.

◆

Discussion

N. D. McMullin (Lenkurt Electric Company, Inc., San Carlos, Calif.): The increased use of commercial telegraph facilities for transmission links in business data processing systems, lends emphasis to the necessary early determination of standardized high-speed data channels and to a standard number of units in the code used for this purpose since this affects both the transmission speed and the accuracy of the data.

The required telegraphic transmission speed will vary because of the considerations pointed out by Mr. Bramhall and, in general, because of the economic cost of the computer and equipment. While the 3-kc standard vehicle voice band can be subdivided into data channels of various widths to permit a choice of transmission speeds, it is probable that eventually two or three standard high-speed data channels will be established.

Mr. Bramhall suggests that a 1,275-cycle-wide half-band may prove to be an economical spectrum unit for data transmission. There is great opportunity in the future for further contribution through the co-ordination of transmission people and the computer manufacturers for determining what speeds are dictated by the economic balance of transmission costs and computer end equipment costs.

Transmission of business data consisting primarily of numerical figures must have high accuracy and, therefore, probably requires a code containing more units than the conventional 5-unit telegraph code to make necessary parity checks. The trend seems to be toward a 7-unit code for business data transmission although this has not yet been clearly indicated.

F. B. Bramhall: It is encouraging to know that a number of industry, user, and professional groups are taking cognizance of the importance of standardization in the data transmission field. The Joint Computer Conference has set up a committee charged with studying the code standardization problem as has also the Radio-Electronics-Television Manufacturers Association. Attention to speed and bandwidth requirements will presumably follow. Our own communication division is at this time organizing a technical committee to work in the computer-communication field. All of these efforts deserve the support of the record communication and computer engineers.

Control System for Integrated Data Processing

PHILIP R. EASTERLIN
MEMBER AIEE

TO MEET the increasing demands from its patrons for automation equipment, the Western Union Telegraph Company has completed the design of an extremely versatile data control system for use with high-speed electronic computers at centralized integrated data-processing (IDP) centers. This system serves as the main artery for data information which can be transmitted simultaneously from cities throughout the country over a Western Union telegraph network. As each unit of classified information is received, it is automatically processed into the proper data category for efficient and high-speed programming by modern computers.

This development opens the automation field to many new applications by providing the vital input and output communication needs for economical use of expensive computing equipment at one centralized point in the patron's scope of operations. This paper describes the new system designated as switching system *203-A*, which has been engineered for initial use in Sylvania Electric's new data-processing center at Camillus, N. Y.

General Application for Sylvania Electric

The data-control system, as applied by Sylvania Electric, is connected by a Western Union private-wire telegraph network to their larger plants located throughout the United States. Over this network, which is also used for administrative messages between any two or more offices data messages are originated at any of the outlying plants and contain three data-routing characters corresponding to a particular data category. Such messages are then transmitted over the wire network into one of a number of input or reperforator receiving positions of the data-control equipment located at the data-processing center. At the reperforator receiving positions, data messages are received on chadless reperforator tape containing the printed characters as well as the punched code holes.

As each complete data message is received at a reperforator receiving position, the tape is fed to an intraoffice transmitter which reads the data-routing characters and automatically switches each message over the proper 5-wire 150-word-per-minute crossoffice circuit to a reperforator at the selected data-storage position. As each message is switched crossoffice, a sequence number check is made to assure that the characters are being punched correctly in the tape by the reperforator at the data-storage position. The crossoffice circuit principle employed is very similar to that used in Western Union's plan *21-A* reperforator switching offices. The reperforator tape at the various data-storage positions is wound on a reel, each reel containing data messages under one or a combined group of categories such as payroll, invoices, inventory, etc.

The classified data on individual tape reels of various data-storage positions are periodically removed, and the reperforator tape is then fed into a converter associated with the electronic computer, in this case Remington Rand's Univac. The converter translates the data information from the reperforator tape into punched card form, and the cards are then used in the conventional manner to transmit onto the magnetic tape of the uniservo units associated with the computer. A reperforator tape-to-magnetic-tape converter is now under development and may be used to by-pass the punched card operation.

In subsequent computer operations, Univac is programmed to process individual data subjects such as hourly, daily, or monthly payroll, invoices, production forecasts, etc., which have been gathered and combined from a number of plants throughout the country. When required, the telegraph network is also used for transmitting information from the data-processing center back to the plant.

A block diagram of the Western Union telegraph network for Sylvania Electric is shown in Fig. 1. Administrative messages can be exchanged directly between any of the offices on the same telegraph circuit. In addition, messages between Sylvania plants on different circuits are relayed by Western Union torn-tape reperforator switching centers located at Camillus, N.Y., Williamsport, Pa., and Salem, Mass. Data messages from any Sylvania plant associated with this network are transmitted into the Camillus IDP center, processed by the data-control equipment, programmed to Univac, and the returnable computer information, when required, is transmitted back to the specified plant through the torn-tape switching centers.[1]

Data Control Equipment

The data control system is housed in five basic types of equipment cabinets. The input or receiving portion of the system is provided by one or more reperforator receiving cabinets which have an upper and lower position. The output of this system where reperforator tapes have been separated into specific categories, is provided by one or more data-storage cabinets, which also have an upper and lower position. Three types of auxiliary control cabinets are used to house the automatic switching controls. These are the automatic switching, transmitter-finder, and autocontrol cabinets. Fig. 2 shows the input and output cabinets, which are similar in appearance, and Fig. 3 shows the group of three auxiliary control cabinets.

For the Sylvania application of the data-control system, the auxiliary control cabinets are packaged to provide a maximum of 50 input or reperforator receiving positions and 50 output or data-storage positions. However, additional input and output positions can be provided, as required, by expanding the switching control equipment.

Preparation of Data Messages at Outlying Offices

Data or administrative messages at various Sylvania plants are originated on Western Union's standard type-*19* sets in the form of a manually prepared perforated tape. This tape passes through a tape transmitter that sends the message characters over carrier or physical line circuits to the distant point and provides a printed monitor copy. On circuits having more than one office, a local selector controls transmission and reception. The selector prevents the transmitter from starting until the cir-

---

Paper 56-655, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 28, 1956; made available for printing May 15, 1956.

Philip R. Easterlin is with the Western Union Telegraph Company, New York, N. Y.

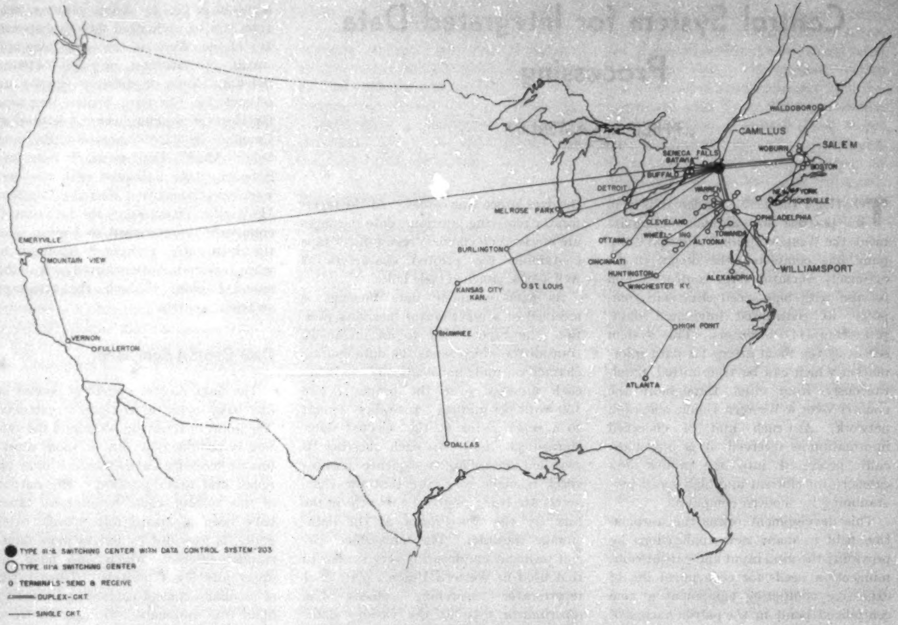

WALDOBORO
CAMILLUS
SALEM
WOBURN
BOSTON
SENECA FALLS
BATAVIA
BUFFALO
DETROIT
NEW YORK
HICKSVILLE
WARREN
MELROSE PARK
CLEVELAND
PHILADELPHIA
EMERYVILLE
MOUNTAIN VIEW
BURLINGTON
OTTAWA
WHEELING
TOWANDA
ALTOONA
WILLIAMSPORT
CINCINNATI
HUNTINGTON
ALEXANDRIA
KANSAS CITY KAN.
ST. LOUIS
WINCHESTER KY.
HIGH POINT
SHAWNEE
VERNON
FULLERTON
ATLANTA
DALLAS
● TYPE III-B SWITCHING CENTER WITH DATA CONTROL SYSTEM-203
○ TYPE III-A SWITCHING CENTER
o TERMINALS-SEND & RECEIVE
DUPLEX-CKT
SINGLE CKT

Fig. 1. Sylvania Electric Products, Inc., private wire system

cuit is idle and allows only the selected stations to copy the transmission.

All data messages destined for the IDP center at Camillus have a message-beginning *Z* character, followed by the routing prefix characters *figures F*, the three data routing characters—the first of which is always the identifying letter *D*—and the message-ending characters of three *carriage returns*. Data messages are either transmitted directly to the data-control system or are relayed to this equipment by means of the torn-tape switching centers. Messages of a administrative nature are preceded by appropriate routing characters and are manually routed to their proper destinations when relayed by the torn-tape switching centers.

Reception of Data Messages

Fig. 4 shows in block form the principles employed in simultaneously receiving data messages from many outlying offices over a telegraph network and in automatically routing each unit of data information into a selected category, as determined by the data-routing characters inserted by the distant or originating office. Messages arrive at the reperforator receiving positions on a printer perforator at 75 words per minute. As the tape is printed and reperforated, it is fed through a newly designed loop-gate intraoffice transmitter.

Upon reading the designated message-beginning character, this transmitter automatically shifts its tape gate to the left and develops a loop of tape while idling and searching for the routing prefix characters embodied some distance from the beginning of the message tape. Upon locating the routing prefix characters, the transmitter stops momentarily until signaled to transmit the three data-routing characters which are used for automatic switching purposes. After transmitting this control information, the tape gate is restored to the normal or right-hand position and, while restoring, the loop of tape is smoothed out to the original starting position, whereupon the complete message tape is then transmitted. Fig. 5 shows two of these transmitters with the tape gate in both positions of operation.

CONNECTION TO AUTOMATIC SWITCHING UNIT

When the message-ending control of three *carriage returns* has been read on the incoming printer perforator, a message-waiting indicator is advanced one step. This indicator is an add-and-subtract unit which indicates the number of unswitched messages by advancing one point on each message-ending signal and subtracting one point for each message switched crossoffice. As the message-waiting indicator is stepped off the zero position, indicating one complete message to be switched, and the intraoffice transmitter has read the routing prefix *figures F*, a request is now made to an automatic switching unit having director equipment capable of serving a group of 12 intraoffice transmitters.

As the automatic switching unit is connected to this particular intraoffice transmitter, the three data-routing characters are received from the transmitter and stored on selector relays located in the automatic switching unit. Contacts on these relays form character-reading

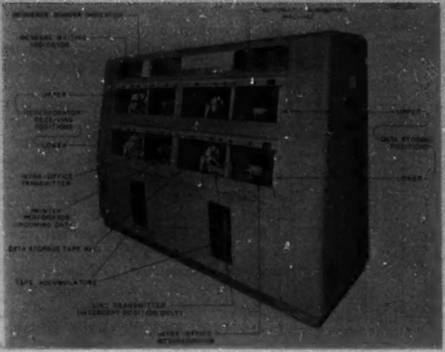

MESSAGE NUMBER INDICATOR
AUTOMATIC NUMBERING MACHINE
MESSAGE WAITING INDICATOR
UPPER
REPERFORATOR RECEIVING POSITIONS
LOWER
INTRA-OFFICE TRANSMITTER
PRINTER PERFORATOR (INCOMING DATA)
DATA STORAGE TAPE REEL
TAPE ACCUMULATORS
SINC TRANSMITTER (INTERCEPT POSITION ONLY)
UPPER
DATA STORAGE POSITIONS
LOWER

Fig. 2. Reperforator receiving and data-storage cabinets

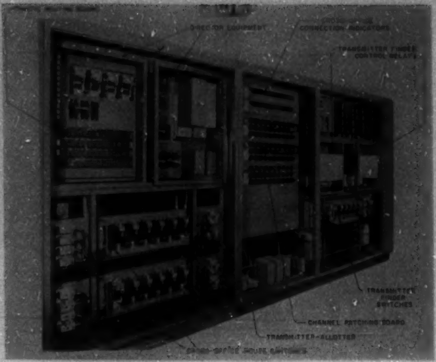

DIRECTOR EQUIPMENT
CROSSOFFICE CONNECTION INDICATORS
TRANSMITTER FINDER CONTROL RELAYS
TRANSMITTER FINDER SWITCHES
CHANNEL PATCHING BOARD
TRANSMITTER-ALLOTTER
CROSS-OFFICE PULSE SWITCHES

Fig. 3. Auxiliary control cabinets

circuits which interpret the three data-routing characters. As these characters, representing any one of a number of data categories, are stored, they cause a rotary-type route switch associated with this intraoffice transmitter to step to the proper stator point corresponding to the selected data category or data-storage position. As this action which requires 3 or 4 seconds is completed, the automatic switching unit disconnects and performs a similar cycle of operation for other intraoffice transmitters having messages to be switched over the crossoffice circuits.

The automatic switching unit, in addition to performing its normal switching functions, also detects administrative messages which have been misrouted into the data-control equipment, or data messages which contain improper or unregistered data-selection characters. For example, in the case of misrouted administrative messages, the director equipment recognizes the absence of *D* as the first of the three data-routing characters and such messages are automatically switched to an intercept data-storage position where they are retransmitted to a receiving-only position for manual handling. However, if *D* is the first of the three data-routing characters, but the following two characters do not represent a valid data category, the tape is held in the intraoffice transmitter and the supervisor is automatically summoned.

ACTUATING TRANSMITTER FINDER

After the automatic switching unit has directed the route switch to the proper stator point, corresponding to the three data-routing characters, the intraoffice transmitter waits for a ground at its individually assigned point on a ring-type transmitter allotter. The transmitter allotter has a rotating wiper which sequentially sweeps all points on the ring every 3.3 seconds and is grounded through the contacts of an allotter-control relay. As an individual transmitter receives an allotment, this relay removes ground from the allotter wiper until this particular crossoffice connection has been completed, thereby preventing two simultaneous connections to the same data-storage position. When an intraoffice transmitter is actuated in this manner, the transmitter finder switch representing the selected data-storage position is initiated, and this switch steps to a marked point corresponding to the intraoffice transmitter seeking connection.

CONNECTION TO THE SELECTED DATA-STORAGE POSITION

As the transmitter-finder switch is properly positioned, relay-control circuits are actuated to complete the crossoffice connection between the intraoffice transmitter at the reperforator receiving position and a 5-wire reperforator located at the selected data-storage position. The 5-wire circuit between these two apparatus units is made through five heavy-duty contacts on a free-running impulse unit operating at 150 words per minute. The five crossoffice intelligence pulse circuits, including a sixth conductor for the step circuit of the intraoffice transmitter, are shown in Fig. 6.

CROSSOFFICE TRANSMISSION

Before crossoffice transmission begins, an automatic numbering machine at the data-storage position—consisting of four rotary switches which are arranged to transmit on a 5-wire basis through the impulse unit—transmits the next consecutive message number characters into the reperforator. Following this number, relay-control circuits connect the impulse unit and the associated intraoffice reperforator to the five crossoffice conductors of the intraoffice transmitter position. Similar control circuits at this position now function to establish proper synchronism with the free-running impulse unit by absorbing the first transmitter step pulse before completing the circuit to the intraoffice transmitter.[2]

The 5-wire reperforator used at the data-storage position, in addition to punching the code holes, is eqpuiped with five read-back transmitte contracts located five characters to the left of the tape punch block that perforates the chad-type holes. Five character operations after a particular character is transmitted into this reperforator, the code holes for this character appear over the read-back contacts. These contacts are connected to the five crossoffice pulse conductors and, as the reperforator punches the code holes of each new character, the read-back contacts close and transmit the fifth previously received character back to the intraoffice transmitter position.

SEQUENCE NUMBER CHECK

Each data message switched crossoffice by the intraoffice transmitter is preceded by a sequence number of about 10 characters which is inserted by an automatic numbering machine located at the distant sending end of the circuit transmitting to

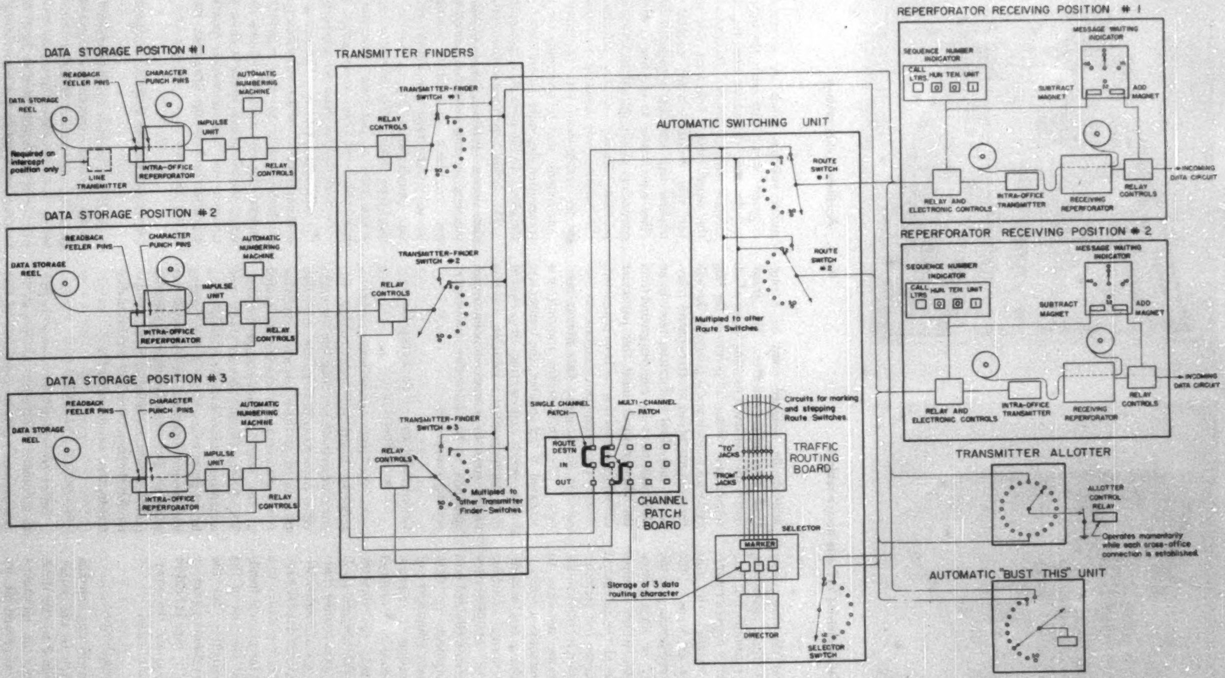

DATA STORAGE POSITION # 1
READBACK FEELER PINS
CHARACTER PUNCH PINS
AUTOMATIC NUMBERING MACHINE
DATA STORAGE REEL
IMPULSE UNIT
Required on intercept position only
LINE TRANSMITTER
INTRA-OFFICE REPERFORATOR
RELAY CONTROLS
DATA STORAGE POSITION # 2
READBACK FEELER PINS
CHARACTER PUNCH PINS
AUTOMATIC NUMBERING MACHINE
DATA STORAGE REEL
IMPULSE UNIT
INTRA-OFFICE REPERFORATOR
RELAY CONTROLS
DATA STORAGE POSITION # 3
READBACK FEELER PINS
CHARACTER PUNCH PINS
AUTOMATIC NUMBERING MACHINE
DATA STORAGE REEL
IMPULSE UNIT
INTRA-OFFICE REPERFORATOR
RELAY CONTROLS
TRANSMITTER FINDERS
TRANSMITTER-FINDER SWITCH # 1
RELAY CONTROLS
TRANSMITTER-FINDER SWITCH # 2
RELAY CONTROLS
TRANSMITTER-FINDER SWITCH # 3
RELAY CONTROLS
Multipled to other Transmitter Finder-Switches
SINGLE CHANNEL PATCH
MULTI-CHANNEL PATCH
ROUTE DESTN
IN
OUT
CHANNEL PATCH BOARD
AUTOMATIC SWITCHING UNIT
ROUTE SWITCH # 1
ROUTE SWITCH # 2
Multipled to other Route Switches
Circuits for marking and stepping Route Switches
"TO" JACKS
"FROM" JACKS
TRAFFIC ROUTING BOARD
SELECTOR
MARKER
Storage of 3 data routing character
DIRECTOR
SELECTOR SWITCH
REPERFORATOR RECEIVING POSITION # 1
SEQUENCE NUMBER INDICATOR
CALL LTRS
HUN
TEN
UNIT
MESSAGE WAITING INDICATOR
SUBTRACT MAGNET
ADD MAGNET
RELAY AND ELECTRONIC CONTROLS
INTRA-OFFICE TRANSMITTER
RECEIVING REPERFORATOR
RELAY CONTROLS
INCOMING DATA CIRCUIT
REPERFORATOR RECEIVING POSITION # 2
SEQUENCE NUMBER INDICATOR
CALL LTRS
HUN
TEN
UNIT
MESSAGE WAITING INDICATOR
SUBTRACT MAGNET
ADD MAGNET
RELAY AND ELECTRONIC CONTROLS
INTRA-OFFICE TRANSMITTER
RECEIVING REPERFORATOR
RELAY CONTROLS
INCOMING DATA CIRCUIT
TRANSMITTER ALLOTTER
ALLOTTER CONTROL RELAY
Operates momentarily while each cross-office connection is established
AUTOMATIC "BUST THIS" UNIT

Fig. 4. Switching theory of data-control system

Vol.
Pt. 1
1956

the data-control system. As these characters are transmitted crossoffice, the read-back contacts on the reperforator at the data-storage position function to transmit back the punched intelligence over the same 5-wire crossoffice conductors and, during this sequence number check, transmission is at half speed, or 75 words per minute. For example, a character is sent from the intraoffice transmitter to the reperforator and, while the code holes are being perforated in the tape, the read-back contacts function to transmit the fifth character previously received back to the intraoffice transmitter position.

These read-back signals are received electronically at the intraoffice transmitter position and are compared on a sequence number indicator having several rotary switches which maintain the sequential numbering pattern. If each of the sequence number characters compares correctly, the intraoffice transmitter is allowed to transmit the next character, etc. This cycle of operation continues until the three identifying characters and a 3-digit number are compared correctly, after which time the comparison circuit is disabled and the remainder of the message is transmitted crossoffice at the normal speed of 150 words per minute. Failure to compare the sequence number actuates automatic test equipment which performs an immediate check on the equipment involved in the crossoffice connection. This action is described later.

When the end-of-message signal, consisting of three *carriage returns*, is transmitted into the data-storage reperforator, the intraoffice transmitter is stopped and the crossoffice circuit is disconnected. This particular data-storage position is now in readiness to receive a crossoffice connection from other intraoffice transmitters having messages prefixed with the same three data-routing characters.

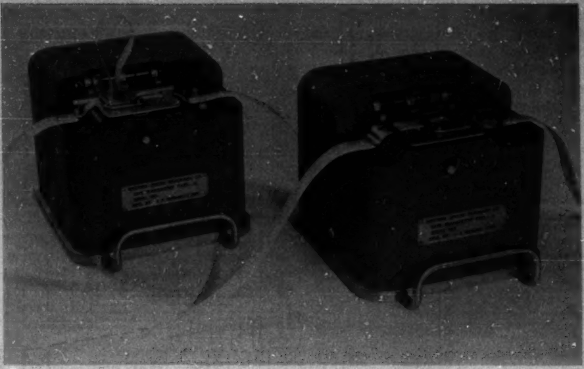


Fig. 5. Intraoffice transmitter

Failure to Compare Sequence Number Correctly

If a wrong comparison occurs on the sequence number check, which can be caused by improperly prepared tape or equipment trouble, the intraoffice transmitter is immediately stopped and the supervisor is summoned. In the meantime, an automatic equipment test is made on this particular crossoffice circuit by an automatic "bust this" unit. This unit functions to cut into any crossoffice circuit where a wrong comparison has occurred and transmits the words *bust this*, followed by the end-of-message signal, into the reperforator at the data-storage position. As these test characters are punched in the reperforator tape, they are transmitted back and compared by the automatic "bust this" unit in a manner similar to the normal sequence number check. If a correct comparison of the test characters is made indicating that the reperforator is functioning properly at the data-storage position, the end-of-message signal following the words *bust this* releases the crossoffice connection and automatically restores the data-storage position to service. Under this condition, the tape is retained in the intraoffice transmitter for the attention of the supervisor. Failure of the test characters to compare properly with the automatic "bust this" unit indicates crossoffice equipment trouble to the data-storage position and the connection between the intraoffice transmitter and reperforator is retained for the attention of the supervisor; in addition, the testing and regulating attendant is alerted.

Multichannel Data-Storage Positions

Provision is made for patching any two or more data-storage positions together to form a multichannel data category to meet conditions where the hourly or daily traffic load exceeds the capacity of one data-storage position. This is performed by a channel patch board located in the autocontrol cabinet. Also, if the capacity of a single data-storage position is unexpectedly exceeded, supervisory attention is automatically directed to this condition when the crossoffice holding time—inability to obtain the desired data-storage connection—exceeds 2 minutes from any one of the intraoffice transmitter positions.

Traffic Routing

The automatic switching unit, which interprets the data-routing characters and automatically positions the route switch associated with each intraoffice transmitter, contains a traffic-routing board which provides complete flexibility in the choice of routing and grouping data messages into any of the data categories. By means of a 2-conductor patch cord, the data-routing characters can be reinterpreted to combine several categories into one data-storage position. For example, it may be desired to group lightly loaded categories on a particular day so that invoice data (DID), production schedules (DPS), inventory reports (DIR), or other specific categories will be received at one data-storage position. Such reassignment or grouping makes it possible to associate the data-storage positions normally assigned to these categories to multichannel purposes on other heavily loaded categories.

LOCAL REDISTRIBUTION OF DATA MESSAGES

Where several different data categories have been grouped into one data-storage position, it is possible to rerun the reperforator tape from this particular tape reel through the system for redistribution on a local basis. For example, it may be desired to accumulate the following data subject matter into one data storage position for redistribution later: payroll hourly (DPH), payroll daily (DPD),

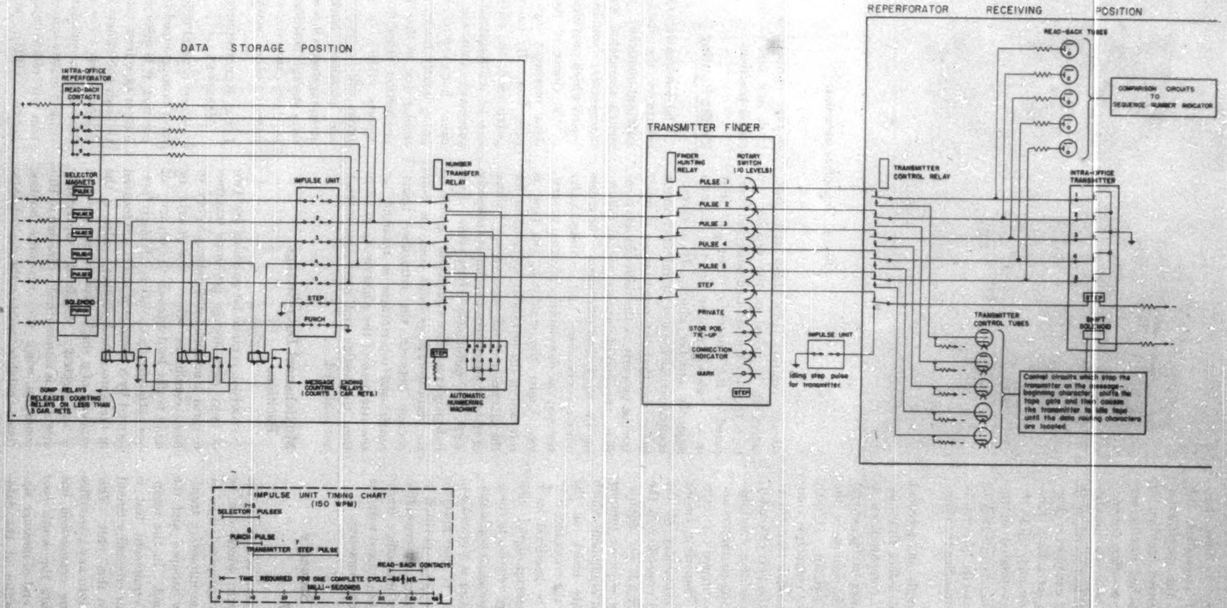

DATA STORAGE POSITION
INTRA-OFFICE REPERFORATOR
READ-BACK CONTACTS
1
2
3
4
5
SELECTOR MAGNETS
PULSE 1
PULSE 2
PULSE 3
PULSE 4
PULSE 5
SOLENOID
PUNCH
IMPULSE UNIT
1
2
3
4
5
STEP
PUNCH
NUMBER TRANSFER RELAY
STEP
AUTOMATIC NUMBERING MACHINE
DUMP RELAYS (RELEASES COUNTING RELAYS ON LESS THAN 3 CAR. RETS.)
MESSAGE ENDING COUNTING RELAYS (COUNTS 3 CAR. RETS.)
TRANSMITTER FINDER
FINDER HUNTING RELAY
ROTARY SWITCH (10 LEVELS)
PULSE 1
PULSE 2
PULSE 3
PULSE 4
PULSE 5
STEP
PRIVATE
STOP FOR TIE-UP
CONNECTION INDICATOR
MARK
STEP
IMPULSE UNIT
Idling step pulse for transmitter
REPERFORATOR RECEIVING POSITION
READ-BACK TUBES
COMPARISON CIRCUITS TO SEQUENCE NUMBER INDICATOR
TRANSMITTER CONTROL RELAY
INTRA-OFFICE TRANSMITTER
1
2
3
4
5
STEP
SHIFT SOLENOID
TRANSMITTER CONTROL TUBES
Control circuits which stop the transmitter on the message-beginning character, shifts the tape gate and then causes the transmitter to idle tape until the data routing characters are located.
IMPULSE UNIT TIMING CHART (150 WPM)
1-5
SELECTOR PULSES
6
PUNCH PULSE
7
TRANSMITTER STEP PULSE
READ-BACK CONTACTS
TIME REQUIRED FOR ONE COMPLETE CYCLE—66⅔ MS.
MILLI-SECONDS
0
10
20
30
40
50
60
66⅔

Fig. 6. Theory of 5-wire crossoffice circuit

payroll weekly (DPW), and payroll monthly (DPM).

By changing the data routing or grouping pattern at the traffic-routing board and rerunning the tape through one or more intraoffice transmitters located at available reperforator receiving positions, the four data categories mentioned in the foregoing may be redistributed to four separate data-storage positions. When data information is reprocessed locally, the normal operating features of the system are retained, including the sequence number check of each data message.

Automatic Switching to Distant Offices

While the data-control system has been designed for meeting the IDP requirements of Western Union's patrons, it also may be used for automatically switching administrative messages to distant offices. This additional operating feature was considered in the design of the data-storage positions, which are arranged to accommodate a line transmitter between the reperforator and the tape reel for retransmission of the perforated tape to distant offices. When used in this manner, the system will function in some respects like Western Union's high-speed reperforator switching centers, now operating in 15 of the larger offices throughout the country.

References

1. A Simplified Telegraph Switching System Plan 111-A, T. S. Pessagno. *Western Union Technical Review*, New York, N. Y., vol. 9, no. 2, Apr. 1955.

2. A New Automatic Message Numbering Machine, J. E. Stebner. *Ibid.*, vol. 8, no. 4, Oct. 1954.

A Video Visual Measuring Set with Sync Pulses

P. T. SPROUL
MEMBER AIEE

Synopsis: This test set is designed to make gain-frequency measurements of a video transmission circuit over the range of 100 kc to 10 mc (megacycles) and present the characteristic on an oscilloscope. The test signal is arranged to include a simulated horizontal synchronizing pulse for use on video circuits equipped with pulse operated devices such as clampers. It is intended for operation where the transmitter and receiver are at separate locations. Provision is made at the receiver for quantitative determination of gain deviation and frequency.

THE OPERATION and maintenance of video transmission circuits requires an easy and rapid means of making gain-frequency measurements. Such information is needed, first, to determine the condition of circuits in service and, second, to assist in making the required adjustments. To meet this need, a video test set has been designed by the Bell Telephone Laboratories which presents the gain-frequency characteristic on an oscilloscope. This set sweeps the video band from 100 kc to a maximum of 10 mc.

A video facility in the telephone plant as furnished to the customer is usually composed of several circuits in tandem, each containing various kinds of equipment. The problem of measuring such facilities has been complicated by the use of clampers and other devices requiring the presence of horizontal synchronizing pulses for proper operation. This transmission-measuring set provides for the optional inclusion of simulated television horizontal synchronizing pulses in the test signal and means in the receiver for measuring in their presence. It was originally intended for the measurement of *TD-2* radio relay circuits on a video-to-video basis, but it can be used on most types of video circuits. In its present form it cannot be used for accurate measurements of coaxial facilities, as the sweeping test signal would interfere with pilot signals used on those systems for gain regulation.

The video test signal used is a sine wave which varies linearly in frequency over the desired video range with a repetition rate of about 70 cps (cycles per second). This rate was chosen from two considerations: First, it is undesirable to sweep at the usual power line frequency rate of 60 cps, as the presence of extraneous power frequency energy cannot be separated from the desired gain-frequency information. With the rate chosen, 60-cycle interference results in multiple traces and is immediately apparent. Second, to obtain maximum resolution, the sweep repetition rate should be as low as possible consistent with excess flicker in the presentation. The receiver switches on alternate sweeps of the transmitter between the signal received from the circuit under test and a single frequency signal of known power, providing a flicker rate of one-half the repetition rate. The rate chosen provides a negligible amount of flicker on the usual *P1* phosphor screen.

The line-synchronizing pulses, having the standard 15,750-cps repetition rate, are switched on when required to operate the clampers of the circuit under test, as already mentioned. They have no function in the test set receiver. In distinction, the oscilloscope is synchronized at 70 cps, the sweep repetition rate, by circuits in the receiver which recover this frequency from the received signal. The wave form of the composite test signal is shown in Fig. 1, where the voltages represent the maximum amplitude available from the transmitter. The sine wave portion of the signal is equivalent to a power of −10 dbm (decibels referred to 1 milliwatt) in a 75-ohm circuit, the nominal impedance of the transmitter and receiver.

System Operation

As shown in the block diagram of the system in Fig. 2, the test set is divided into a transmitter and a receiver for use at the ends of the facility being tested. The transmitter generates the test signal and consists of the video sweep generator and the synchronizing pulse generator. These are assembled with the necessary power supply in a portable unit, as shown in Fig. 3. The unit may be removed from the carrying case and may be rack-mounted, if desired. A video attenuator is provided to control the composite output voltage over a maximum range of 20 db (decibels) in 1-db steps. A sync-amplitude control permits varying the amplitude of the synchronizing pulse.

Paper 56-673, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 29, 1956; made available for printing May 15, 1956.

P. T. Sproul is with Bell Telephone Laboratories, Inc., Murray Hill, N. J.

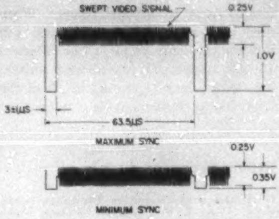


Fig. 1. Test signal wave form

Thus, the per cent of synchronizing pulse may be set as desired.

The receiver provides for detecting the sweeping video test signal and presenting the envelope of the sine wave portion on the scope with gain and frequency calibration information. The amplitude modulation envelope of the sine wave signal results from gain deviations versus frequency of the circuit under test. These circuits and their power supply are assembled into the portable unit shown in Fig. 4. A video attenuator permits varying sensitivity of the receiver over a 20-db range. Controls are also provided for positioning the calibration trace, for setting the frequency marker, and for adjusting the receiver presentation to be in synchronism with the transmitter repetition rate.

When the received test signal does not include the sync pulse, it is applied directly to the amplifier-detector. A filter is used at its output to separate the wanted envelope information from the original signal. This filter is of the constant resistance type having a nominal cutoff of 25 kc, and it has good transient response. The cutoff frequency of the filter represents a balance between achieving as low a frequency of measurement as possible and obtaining adequate resolution in the measurement.

It is obvious that measurements cannot be made below the cutoff of the filter because the wanted envelope information cannot be separated from the original signal information. In addition, the transient response of the filter makes the practical minimum frequency of measurement somewhat higher than the nominal filter cutoff. A value of 100 kc is reasonable for the filter used here.

The filter cutoff must be high enough to obtain reasonable resolution. Allowing about 1.5 milliseconds for receiver switching, the useful duration of the transmitter sweep is about 12.5 milliseconds. When sweeping the entire 10-mc band, this corresponds to a rate of change of sweep frequency of approximately 800 cps per microsecond. Assume that the characteristic being measured is a sine wave over the entire video band, equivalent to a single echo. The maximum envelope frequency that the filter will transmit without attenuation is 25 kc or a sine wave having a period of 40 microseconds. With the sweep speed mentioned, the sweep will change in frequency 32 kc in 40 microseconds; therefore this is the finest ripple that will be presented at its true amplitude. Such resolution is adequate for the usual video circuits.

The presence of the sync pulse in the signal causes improper operation of the detector in the receiver. To detect the envelope of the sine wave signal in its presence, the incoming composite signal is passed through a high-pass filter. This filter removes most of the energy in the sync pulse, permitting the detector to operate normally. It is possible to separate these two signals on a time basis, but the additional complexity was not justified for this application. The characteristic of the high-pass filter is shown in Fig. 5. The cutoff frequency was determined experimentally to be as low as possible consistent with proper detector operation. As the figure shows, the nominal value is of the order of 250 kc. The use of this filter increases the lowest frequency at which measurements can be made to about 500 kc. Because of this, provision is made for switching it out when sync pulses in the test signal are not required. It should be noted that the 25-kc filter at the output of the detector is switched out when the high-pass filter us used. This is necessary because of its poor transient response in the presence of the residual sync pulse information. A filter with a higher frequency cutoff was not necessary because of the nominal 200-kc cutoff of the oscilloscope used.

The input to the receiver is switched on alternate sweeps of the transmitter between the incoming signal and a single frequency signal of known power whose level can be varied by means of a calibrated attenuator. This gives the receiver presentation shown in Fig. 6. The determination of gain deviations can be made by superimposing the calibration trace on the line trace at the points of interest and reading the relative gain from the calibration attenuator. This makes the results independent of

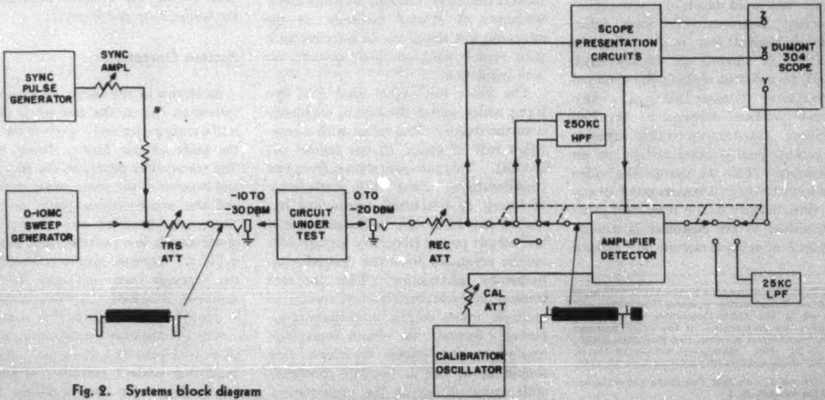


Fig. 2. Systems block diagram

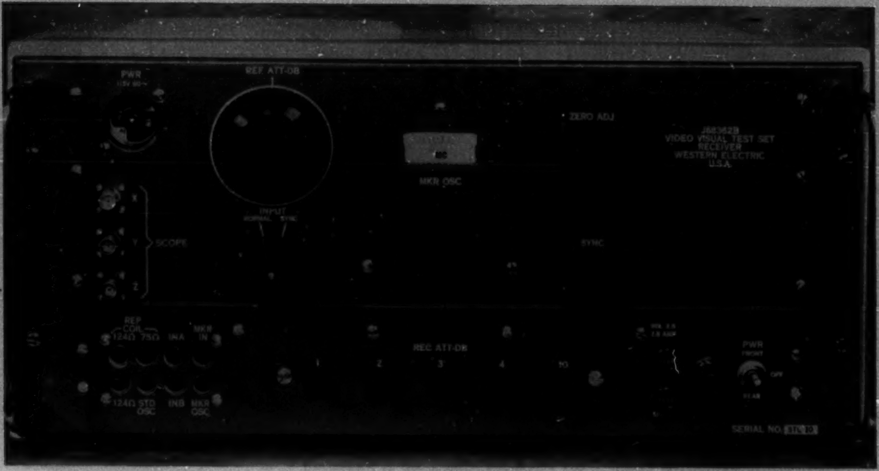

PWR
REF ATT-DB
ZERO ADJ
J68362B
VIDEO VISUAL TEST SET
RECEIVER
WESTERN ELECTRIC
U.S.A.
MKR OSC
INPUT
NORMAL
SYNC
X
Y
Z
SCOPE
SYNC
REP COIL
124Ω
75Ω
INA
MKR IN
REC ATT-DB
1
2
3
4
10
PWR
FRONT
OFF
REAR
124Ω
STD OSC
INB
MKR OSC
SERIAL NO.

Fig. 3. Transmitter unit

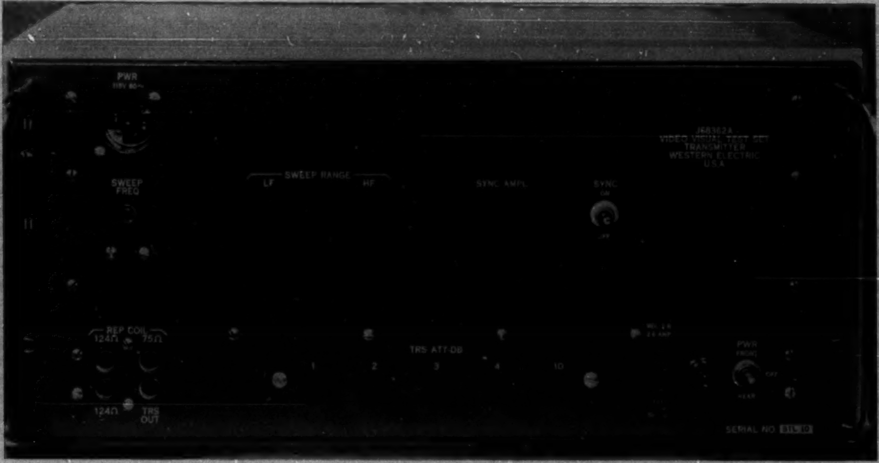

PWR
J68362A
VIDEO VISUAL TEST SET
TRANSMITTER
WESTERN ELECTRIC
U.S.A.
SWEEP FREQ
LF
SWEEP RANGE
HF
SYNC AMPL
SYNC
ON
OFF
REP COIL
124Ω
75Ω
TRS ATT-DB
1
2
3
4
10
PWR
FRONT
OFF
REAR
124Ω
TRS OUT
SERIAL NO.

Fig. 4. Receiver unit

amplifier-detector or oscilloscope gain settings and dependent only on the attenuator accuracy and the flatness of the transmitter and receiver gain frequency characteristic. D-c transmission is provided between the detector and the d-c oscilloscope to permit independent motion of the two traces when manipulating their respective attenuators, which makes operation of the set simpler. The frequency marker may be set to any frequency desired over the 100-kc to 100-mc range.

The set is designed so that the transmitter and receiver can be operated at the ends of a 1-way circuit. In addition to the necessity of providing for synchronizing the receiver presentation with the transmitter, this places stringent requirements on the gain-frequency characteristic of each unit. It is essential that the transmitter work with any receiver in the system, and vice versa. The only practical way to achieve this is to have each unit as flat as possible.

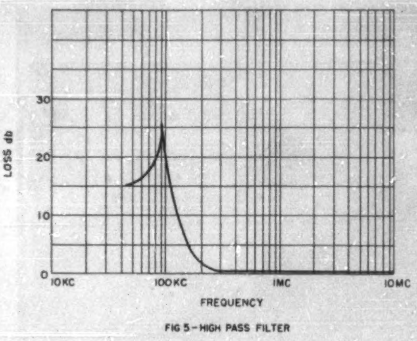


Fig. 5. High-pass filter characteristic

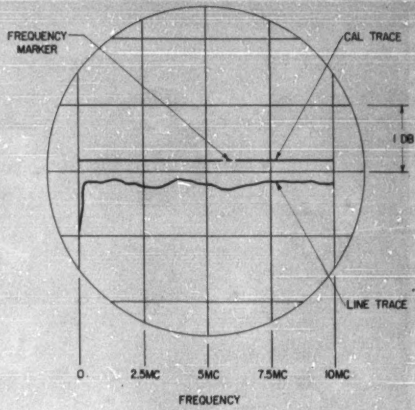


Fig. 6. Oscilloscope presentation

For this, the receiver video circuits were conservatively designed and can be aligned, using a thermocouple as a standard, to within ±0.02 db over the 100-kc to 10-mc band. The receiver, after alignment, is used to adjust the transmitter. A typical set can be adjusted on a back-to-back basis to be flat to within ±0.03 db over the entire 10-mc band. The residual deviation of the set has a period of the order of 5 mc. It is of interest to the television engineer that this is equivalent to a 0.2-microsecond echo which is 50 db below the signal. In discussing the influence of echos on a television picture, Mertz[1] reports subjective data showing that an echo having this delay is barely perceptible at about 25 db below signal. There is therefore considerable margin in the performance of this test set. The accurate measurement of finer structure deviations is readily accomplished without being masked by test set deviations.

Receiver synchronization, frequency marker, and scope presentation circuits are controlled from the incoming signal. These operate to keep the receiver presentation in synchronization with the transmitter and to provide the continuously variable frequency marker appearing on the calibration trace.

Transmitter

The transmitter is shown in block diagram form in Fig. 7. The swept video signal is the difference frequency between a fixed 80-mc oscillator and a variable 70- to 80-mc oscillator and is generated in a silicon diode modulator. The variable 70- to 80-mc oscillator is controlled by a 70-cycle saw-tooth generator of conventional design. The panel controls shown make possible adjustment of the video sweep range of the transmitter. The low-frequency adjustment (LF ADJ) sets a voltage to which the positive excursion of the saw-tooth wave is clamped. This corresponds to the low-frequency end of the video output. The high-frequency adjustment (HF ADJ) sets the amplitude of the saw tooth and determines the high-frequency end of the video output. The provision for clamping the positive peak of the saw tooth provides for changing the high-frequency end of the sweep without affecting the low-frequency end setting, thus simplifying adjustment.

A simplified schematic of the video sweep generator is shown in Fig. 8. The 70- to 80- mc sweep oscillator is of the reactance tube modulated phase shift type as described by Dennis and Felch.[2] A high-transconductance pentode is used for the reactance tube and a triode for the oscillator. Three of the elements of the transmission line are adjustable as shown. They are set to give the desired sweep range with maximum flatness with the use of a fixed amplitude of saw-tooth voltage on the reactance tube grid. The buffer amplifier feeds an equivalent T transformer to match it to the low-impedance modulator. This is adjusted for maximum flatness over the 70- to 80-mc range of the oscillator with a suitable detector connected to the output of the transformer. The 80-mc oscillator is crystal-controlled, thus eliminating the problem of measuring frequency during alignment. Its buffer amplifier drives a transformer with a tuned primary. The secondary of this transformer is balanced to ground and connects to the balanced input of the silicon varistor modulator. The 70- to 80-mc oscillator provides the high level input to the modulator, and the clipping which results aids in reducing output amplitude variations as the frequency is swept. The video output power is determined by the 80-mc input voltage to the modulator which is set by a capacity potentiometer in the grid of the 80-mc buffer. Filters are provided as indicated to suppress the high-frequency energy in the video output. As the impedance of the modulator is of the order of 100 ohms, a matching pad is provided to give the desired 75-ohm output. This also acts to improve the output impedance of the transmitter.

The sync pulse generator consists of a 15,750-cps sine wave oscillator triggering a multivibrator to generate the desired sync pulse. This pulse is applied in a positive direction across the cathode of the 80-mc buffer in the sweep generator to turn off the sine wave energy during the sync pulse interval. It is also applied in a negative direction through a large resistor across the output of the video sweep generator to add the pulse during the blanking interval. Wave forms are shown in Fig. 7.

A repeating coil and equalizer is provided to permit measuring the 124-ohm balanced circuits used extensively in the telephone plant. They have an insertion loss of about 0.6 db and are flat over the 100-kc to 10-mc range to within ±0.02 db.

Vol,
Pt. 1
1956

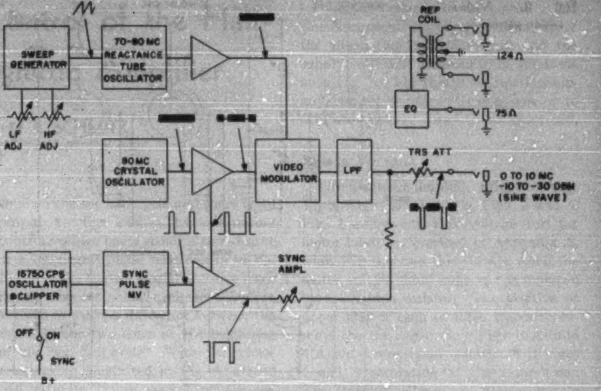


Fig. 7. Transmitter block diagram

Receiver

The block diagram of the receiver is shown in Fig. 9. Two inputs to the receiver are provided. The signal from the equipment under test is normally applied to the IN *A* jack. The following video attenuator provides for adjustment of the incoming signal power to the nominal receiver operating value of −20 dbm. The incoming signal is bridged at this point with a 1-stage video amplifier to feed the receiver presentation circuits. The high-pass filter follows. The second input to the receiver, the IN *B* jack, is normally connected to a 5-mc oscillator whose power output is variable in known amounts. The power at the CAL OSC jack is initially set up to be −20 dbm when the calibration attenuator CAL ATT is set to zero dial reading. The attenuator then provides for varying the power of the calibration signal ±4.4 db in 0.2 db steps.

Details of the video switch and amplifier-detector circuits are shown in Fig. 10. The two inputs are terminated in cable impedance using 75-ohm resistors. The switch transfers the grid of the video amplifier stage from one terminating resistor to the other. This is possible because the specially designed switch has a parasitic capacity to ground of only 7 micromicrofarads. It is of the reed type with mercury contacts and, as used in this set, has an average operating time of about 1.5 milliseconds. The video amplifier consists of one stage using a high-transconductance pentode with a simple plate-peaking circuit. It drives an envelope detector using two silicon diodes. This type of detector was chosen to minimize errors arising from the effect of harmonics generated either in the transmitter or in the circuit under test. Harmonics 25 db below the fundamental will cause an error of only 0.01 db in the output of this detector. As used, the output is balanced, and the following amplifier stage not only provides for d-c amplification but also translates to an unbalanced output to feed the oscilloscope. The input to the oscilloscope must, of course, be with reference to ground. The translation from the plate voltage of the d-c amplifier to ground is accomplished by using a divider and a 90-volt bucking battery of the portable radio type. Although the use of a battery for this purpose might seem undesirable, the severe requirements on ripple and voltage stability of the alternative regulated negative supply indicated a substantial saving in size and complexity by the use of the battery. The battery units have essentially shelf life as their load is the 2-megohm input impedance of the scope. The SCOPE POS control provides for exact balance of the output voltage to ground and is used for vertical positioning of the scope picture.

The receiver presentation circuits fed from the video bridging amplifier perform the functions of providing the frequency marker and generating the scope display. One output from the bridging amplifier is introduced into a detector together with the output of a video marker oscillator. The signal generated by this detector in the region of zero beat is amplified and shaped for applying to the *Z* axis or intensity grid of the oscilloscope. The marker generated is a blanked spot at the frequency of the video marker oscillator. In the center of this blank spot is a flickering dot which is the point at which the sweeping input goes through zero beat with the fixed marker oscillator. It flickers because of the random phase at which zero beat occurs. Considerable amplification and clipping action is provided in the pulse-shaping circuits to make the frequency marker independent of signal voltage input over a wide range.

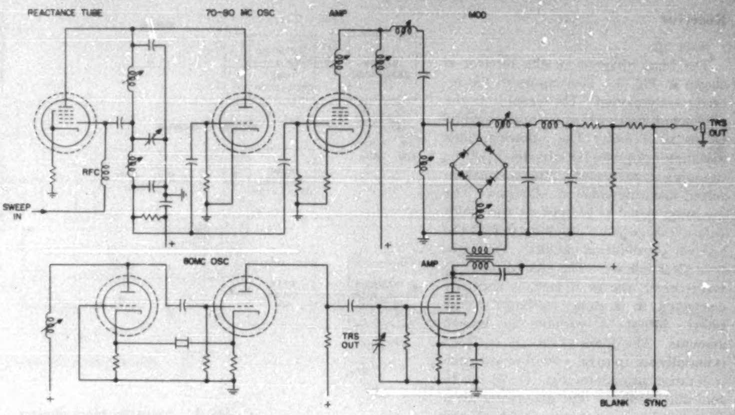
The marker oscillator is of the beat oscillator type in order to obtain the entire range of 0 to 10 mc on one dial scale. It operates by beating a crystal-controlled 80-mc oscillator against a 70- to 80-mc variable oscillator in a silicon diode modulator. The output is at one milliwatt and the marker circuit may be energized alternatively by any oscillator providing this power. Relatively good frequency accuracy is achieved by the use of high-quality components and rigid mechanical design. An accuracy of 100 kc may be expected.

The second output from the bridging amplifier feeds a resistance-capacitance low-pass filter and a detector, forming a simple discriminator. Its output is a crude saw tooth, but the flyback portion is an accurate indication of the time at which the transmitter retraces. This is used to synchronize a blocking oscillator in a circuit similar to the vertical sync circuit of a television receiver. The output pulse of the blocking oscillator controls the receiver presentation. It triggers the switch multivibrator which is of the bistable type, switching at half the transmitter repetition rate. This operates the video switch, causing the input to the amplifier detector to be connected to the incoming signal or calibration signal on successive transmitter sweeps. The switch multivibrator also provides a square wave for the *Z* axis mixer for the purpose of presenting the frequency marker only on the calibration trace. This avoids obscuring the gain-frequency information at the frequency of interest.

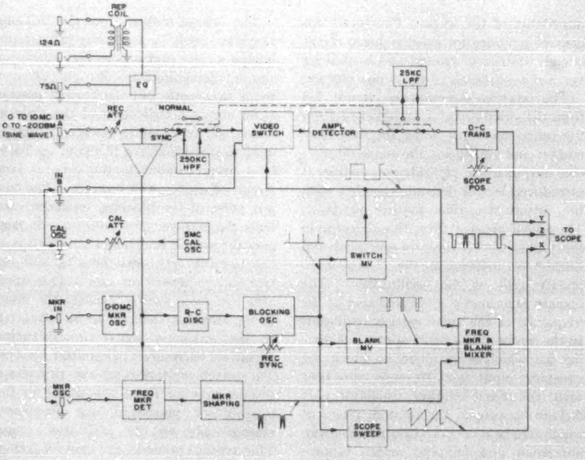
The blocking oscillator pulse performs two other functions. It triggers a scope sweep generator and a blanking multivibrator. The scope sweep generates a saw tooth which is applied to the horizontal input to the scope. The blanking multivibrator is of the monostable type and generates a pulse about 2 milliseconds long. This is used to blank the scope during the operate time of the video relay. The three signals required for the intensity grid of the scope are combined in the mixer.

A repeating coil and equalizer are also provided in this unit to permit measuring 124 ohm balanced video circuits.

Fig. 8. Video sweep generator



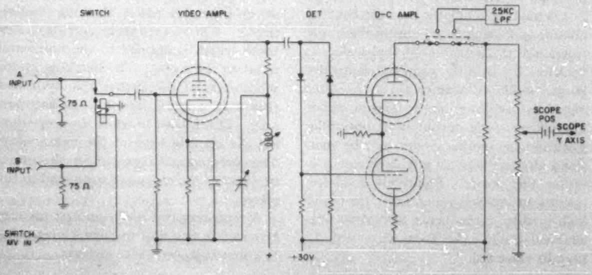
REACTANCE TUBE
70-90 MC OSC
AMP
MOD
RFC
SWEEP IN
80MC OSC
AMP
TRS OUT
TRS OUT
BLANK
SYNC



REP COIL
124Ω
75Ω
EQ
0 TO 10MC IN
0 TO -20DBM
A
(SINE WAVE)
REC ATT
NORMAL
SYNC
250KC HPF
VIDEO SWITCH
AMPL DETECTOR
25KC LPF
D-C TRANS
SCOPE POS
IN B
CAL OSC
CAL ATT
5MC CAL OSC
SWITCH MV
Y
Z
X
TO SCOPE
MKR IN
0-10MC MKR OSC
R-C DISC
BLOCKING OSC
REC SYNC
BLANK MV
FREQ MKR & BLANK MIXER
MKR OSC
FREQ MKR DET
MKR SHAPING
SCOPE SWEEP

Fig. 9 (left). Receiver block diagram

Fig. 10 (below left). Amplifier-detector circuit



SWITCH
VIDEO AMPL
DET
D-C AMPL
25KC LPF
A INPUT
75 Ω
B INPUT
75 Ω
SWITCH MV IN
SCOPE POS
SCOPE Y AXIS
+30V

Summary

The test set described has been found very useful in the measurement of video circuit transmission. For straightaway use, the accuracy is of the order of ±0.05 db. It can be used for the measurement of passive devices having a loss of as much as 10 db with full sensitivity.

References

1. Influence of Echos on Television Transmissions, P. Mertz. *Journal*, Society of Motion Picture and Television Engineers, New York, N. Y., vol. 60, May 1953, p. 572.

2. Reactance Tube Modulation of Phase Shift Oscillators, F. R. Dennis, E. P. Felch. *Bell System Technical Journal*, New York, N. Y., vol. 28, Oct. 1949, p. 601.

Alternating Current Control of the Half-Wave Bridge Magnetic Amplifier

T. BERNSTEIN
ASSOCIATE MEMBER AIEE

N. L. SCHMITZ
ASSOCIATE MEMBER AIEE

RECENT developments in transistors indicate great promise for a high-gain, high-power-output servo amplifier consisting of a transistor preamplifier in combination with a fast-response half-wave bridge magnetic power amplifier. Existing literature on this type of magnetic amplifier does not discuss certain problems inherent in applying this device as an output stage fed by an a-c preamplifier.[1,2] Among these are the problems of reduced output as a result of large control signals, and variation in operation caused by different types of load.

This paper considers various wave forms originating in the magnetic amplifier with the aim of providing a better understanding both of normal operation and of overdriven operation during which output is reduced as a result of large alternating voltage control signals. The effects of various types of load upon firing angles are studied both through theoretical analysis and through investigation of experimentally determined wave forms.

Operation With Alternating Voltage Control

The half-wave bridge magnetic amplifier considered in this paper is shown in Fig. 1. It uses two identically wound cores, each of which is wound with a control winding, two load windings, and a reset winding. The instantaneous polarity of V_S is shown in Fig. 1 for the set half-cycle. Reset windings, N_{R1} and N_{R2}, have their circuits so adjusted that in the absence of a control signal, the cores are reset to equal flux levels before the beginning of each set half-cycle. Both cores then reach saturation at the same time on the set half-cycle and there is no output to the load. The alternating control voltage bears either a zero-degree or a 180-degree phase relationship to V_S, Fig. 1. The control magnetomotive force during the reset half-cycle is therefore such as to cause a difference between the fluxes of the two cores at the beginning of the set half-cycle. With the phase relationship illustrated in Fig. 1, core 2 will saturate before core 1 on the set half-cycle. After the saturation of core 2, and until core 1 reaches saturation, there will be an output to the load. The load is then shunted by the windings on the saturated cores. If the phase of the control voltage is shifted 180 degrees, core 1 reaches saturation before core 2 and thus an output voltage of opposite polarity is obtained.

As the control voltage V_C of Fig. 1 is increased, the output voltage across the load increases until V_C reaches a value such that core 2 remains saturated for the entire set half-cycle. If circumstances warrant, the design of the circuit can be such as to prevent core 1 from saturating during any portion of the set half-cycle. This is a condition for maximum output. As the control voltage is increased, the output voltage tends to limit at its maximum value.

A further increase in control voltage causes the output voltage to decrease. This effect is shown in Fig. 2 for the magnetic amplifier described in Appendix I. This reduction of output voltage with large alternating control voltage is a serious problem when an electronic or transistor preamplifier is used, since large error signals to the input of the preamplifier cause a very large control voltage to appear at the control input of the magnetic amplifier. This large signal in turn causes the output voltage to drop well below its maximum value. When direct control voltage is used, there is no reduction of the output voltage for large control signals.

Explanation of Wave Shapes

To provide a more detailed understanding of the operation of this type of magnetic amplifier, wave forms are shown for various portions of the circuit. For this discussion, the phase relationship of the alternating control voltage and the supply voltage is assumed to be the same as in Fig. 1. For this phase relationship, any control signal will cause core 2 to saturate before core 1, as described.

BRIDGE VOLTAGE

The magnetic-amplifier bridge voltage V_{12} of Fig. 1, is shown in Figs. 3 and 4. Fig. 3 represents this voltage for the motor load, as described in Appendix A, while Fig. 4 shows this voltage when the motor and its shunt capacitor is replaced by a 390-ohm resistor. The positive or upper half of each of these traces represents the set half-cycle. Figs. 3(A) and 4(A) show zero signal operation with cores 1 and 2 saturating at the same time. The bridge voltage is approximately equal to the supply voltage V_S until both cores saturate. This is true since the voltage drop through R_S caused by the exciting current is negligible. After both cores saturate, there is an abrupt drop in bridge voltage because the impedance of the saturated cores is much less than R_S. Figs. 3(B) and 4(B) show the bridge voltage when a small control signal (less than that required for maximum output) is applied. Until core 2 saturates, the bridge voltage is approximately equal to the supply voltage. After core 2 saturates, the bridge voltage is reduced as a result of the increased voltage drop through R_S caused by the load current. When core 1 saturates later in the set half-cycle, the bridge voltage drops to a low value as a result of the low saturated impedance of both cores. Fig 3(C)shows that with the motor load, core 2 is saturated for the entire set half-cycle, while core 1 never reaches saturation. For the resistance load, Fig. 4(C) shows that at maximum output core 2 is saturated for the entire set half-cycle, whereas the time of saturation of core 1 (the second core to saturate) is almost unchanged from the quiescent condition. This phenomenon is explained in the section "Determination of Firing Angles."

Figs. 3(D) and 4(D) show the bridge voltage after the load voltage has begun to decrease as a result of excessive alternating voltage control. The figures show that the sharp drop in the bridge voltage at the trailing edge of the set half-cycle advances as the control voltage is increased. This advance results in less output voltage. Since the sharp drop is a result of the saturation of core 1, reduction in output must stem from early saturation of core 1.

Paper 56-626, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted October 17, 1955; made available for printing April 12, 1956.

T. BERNSTEIN is with General Motors Corporation, Milwaukee, Wis., and N. L. SCHMITZ is with the University of Wisconsin, Madison, Wis.

This work is an extension of studies conducted by Mr. Bernstein as a thesis in partial fulfillment of the requirements of the degree of Master of Science at the University of Wisconsin.

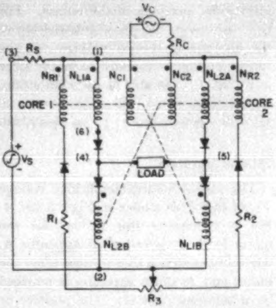


Fig. 1. Half-wave bridge magnetic amplifier circuit

Rectifier Voltage

The voltage V_{64} of Fig. 1 is shown in Fig. 5. Fig. 5(A) through 5(D) represent operation with control signals ranging from zero (quiescent condition) to a value large enough to cause output drop-off; see Fig. 5(D).

During the reset half-cycle, the rectifier voltage is such that point 4 is positive with respect to point 6. This is a nonconducting condition. As the control signal is increased, the voltage V_{46} can be seen to decrease. At first glance, it might be suspected that voltages induced in N_{L1A} and N_{L2B} by windings N_{C1} and N_{C2} respectively should be equal and opposite and have no effect in reducing V_{64} during the reset half-cycle. However, such is not the case, because of of the saturated condition of core 2 at the beginning of the reset half-cycle. Since N_{C1} tends to drive core 1 out of saturation, winding N_{L1A} can produce a back voltage opposing V_{21} and reducing V_{46}. Winding N_{C2} however, tends to drive core 2 further into saturation and therefore can produce no flux change in core 2, and consequently no voltage in winding N_{L2B} tending to aid V_{21}.

During the set half-cycle, the voltage V_{64} is normally only slightly positive because the rectifier is conducting in the positive direction. However, with an alternating control voltage sufficiently large to cause a reduction in output voltage, it is noted experimentally (as in Fig. 5(D)) that point 4 becomes positive with respect to point 6 during the set half-cycle. This potential difference can exist because of the high reverse resistance of the rectifier. The fact that V_{16} can and does exceed V_{14} during the set half-cycle plays an important part in explaining output drop-off in the following section.

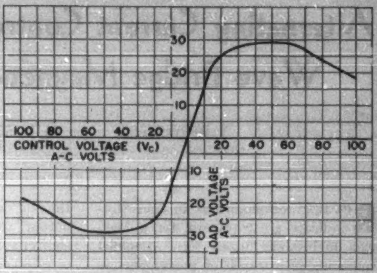


Fig. 2. Transfer characteristic with alternating voltage control

Source Current

The voltage across R_S, V_{31}, is important since it can also be used to determine the firing angle of each core. Fig. 6 shows this voltage when the motor load is used. This voltage represents the current obtained from V_S. Fig. 6(A) is the quiescent condition where the current rises sharply as both cores saturate during the set half-cycle. With a small control signal, Fig. 6(B) shows the sharp increase of current when core 2 saturates, and also the second sharp increase when core 1 saturates and shunts the load. For the maximum output, the wave shape of the current is that of Fig. 6(C). With excessive control signal, the wave shape of Fig. 6(D) is obtained. Fig. 6(D) shows that core 1 now saturates earlier during the set half-cycle and thus reduces output.

Load Voltage

Fig. 7(A) is the timing wave and represents a 115-volt rms, 400-cps (cycle-per-second) wave. Figs. 7(B), 7(C), and 7(D) show the output voltage V_{64} across the motor load for the conditions of output before maximum, at maximum, and after maximum output respectively.

Reduced Output With Large Alternating Control Voltage

The foregoing descriptions of wave forms reveal that output drop-off with large control signals of the phase shown in Fig. 1, occurs because of the earlier saturation of core 1, the second core to saturate. For an understanding of the problem it is necessary to determine why core 1 saturates before the end of the set half-cycle.

For the control signal phase relationship shown in Fig. 1, the control signal tends to oppose the saturation of core 1 on the reset half-cycle. The control signal tends to aid the saturation of core 1 on the set half-cycle. Consequently the early saturation of core 1 as a result of large alternating control voltage seems likely to be caused by some action of the control voltage during the set half-cycle. Experiments confirm this view; they show that if the control voltage is blocked by a rectifier during the set half-cycle, early saturation of core 1 and load voltage drop-off with a large control signal does not occur.

The problem of output drop-off with high alternating control voltage is now isolated. The drop-off occurs because the core which should stay desaturated during the entire set half-cycle saturates before the end of this half-cycle. This type of saturation results from the action of the control on the set half-cycle.

Cause of Early Saturation in Core 1

It has been shown that core 1 saturates as a result of control circuit action during the set half-cycle. Consequently, causes of this saturation are conveniently discussed in terms of the circuit of Fig. 8, which applies during that part of the set half-cycle after core 2 has saturated. This circuit is obtained from that of Fig. 1 by regarding both the saturated impedance of windings N_{L2A} and N_{L2B} and the forward resistance of their associated rectifiers as being negligibly small.

There is a net forward voltage across the load circuit rectifiers of core 1 as long as the supply voltage V_{L1} is greater than the load-winding induced voltage. If the induced voltage in the load winding becomes larger than V_{L1}, then the load-winding rectifier has reverse voltage applied and acts as an open circuit. Thus, when the load-winding rectifier can conduct, the load-winding determines the flux change on the set half-cycle. When current in the load winding is blocked, the control windings determine the flux change. Until the control voltage V_{C1} becomes sufficiently large to induce a voltage in the load windings of core 1 greater than V_{L1}, the voltage V_{L1} determines the flux change in core 1 on the set

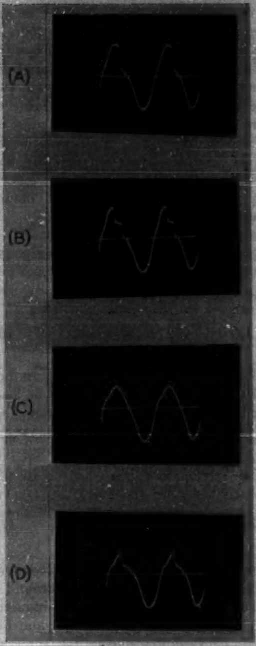


Fig. 3. Bridge voltage v_{12}

A—Quiescent
B—Before maximum output
C—During maximum output
D—After drop-off, with excessive control signal

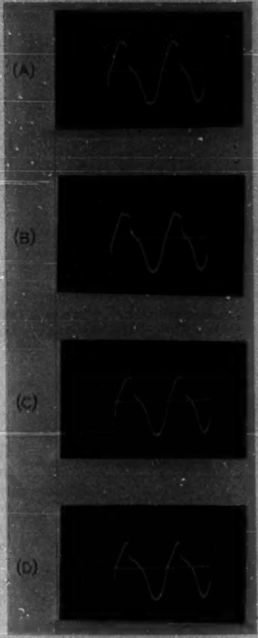


Fig. 4. Bridge voltage v_{12} with 390-ohm resistive load substituted for motor load

A—Quiescent
B—Before maximum output
C—During maximum output
D—After drop-off, with excessive control signal

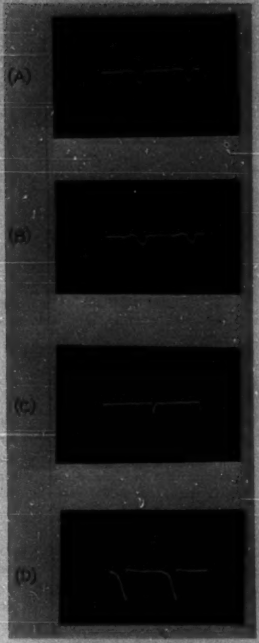


Fig. 5. Rectifier voltage v_{34}

A—Quiescent
B—Before maximum output
C—During maximum output
D—After drop-off, with excessive control signal

half-cycle. This flux change is not sufficient to bring the core to saturation. When the voltage induced in the load windings N_{L1A} and N_{L1B} becomes greater than V_{L1}, there is reverse voltage on the load-winding rectifiers, and the control voltage determines the flux change. This flux change is now greater than that due to V_{L1}, and core 1 can be driven into saturation earlier than it would be with V_{L1} acting alone. The occurrence of the reverse rectifier voltage for large control signal during the set half-cycle is shown in Fig. 5(D).

Prevention of Output Drop-Off

Thus far, the effect of the reset winding in determining the load voltage drop-off has been ignored. These windings usually have fewer turns than the control windings so that the input impedance looking into the control windings is maximized.[2] Consequently the control voltage produces insufficient voltage in the reset winding to overcome V_{L1} and cause reset current during the set half-cycle.

It is now suggested that by adjusting the reset circuit, the control voltage can be made to cause reset current before this voltage can cause the load rectifiers to block. In this case, the reflected current in the control circuit will cause a voltage drop in R_C acting to limit V_{C1} and prevent load voltage drop-off. One method of achieving the desired reset circuit behavior is to supply the reset circuit from a separate and lower voltage source and then reduce the reset resistors as necessary to maintain the original reset action. This, however, reduces both input impedance and gain.[3] Similar disadvantages apply if the limiting is achieved by increasing the number of reset winding turns. It seems most practical to prevent drop-off by inserting a limiter circuit in the preamplifier.

Determination of Firing Angles

In the design of an output stage, the primary objective is to obtain maximum useful output with a minimum of apparatus. Considerations of output to input power ratio are usually secondary as long as the input power requirement is not excessive. Particularly in the case of a motor load, a knowledge of the relationship between firing angles of the two cores is desirable in order to achieve a combination of circuit parameters yielding firing angles which will provide a maximum fundamental component in the harmonically rich output voltage supplied to the motor.

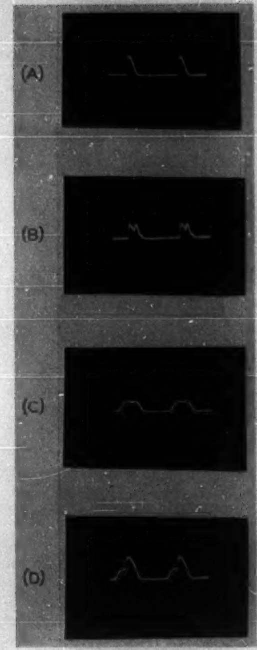


Fig. 6. Current in R_B, v_{21}

A—Quiescent
B—Before maximum output
C—During maximum output
D—After drop-off, with excessive control signal

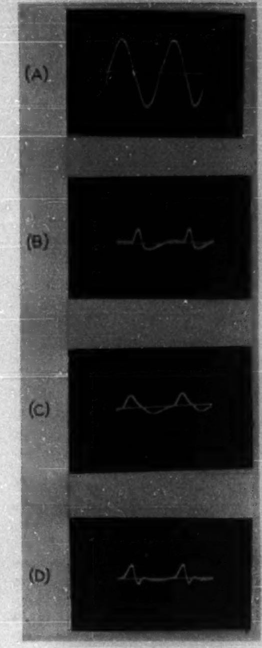


Fig. 7. Motor load voltage v_{54}

A—115-volt rms, 400 cps, source voltage $v_s = v_{32}$
B—Before maximum output
C—During maximum output
D—After drop-off, with excessive control signal

Determination of First Firing Angle

As shown in Fig. 1, the polarity of V_C is such that during the reset half-cycle (point 3 negative) the flux in core 2 is increased relative to the flux in core 1. Core 2 will then fire first, at some angle θ_2, determined by the difference $\Delta\phi_2$ between its flux level at the beginning of the set half-cycle and the flux level for saturation. It follows that the maximum change in flux linkages which can occur with an N_L-turn loading winding on core 2 in the time interval between $t = 0$ (the beginning of the set half-cycle) and $t = t_2$ (the time at which core 2 saturates) is $N_L\Delta\phi_2$. This fixes the time t_2 in accordance with equation 1.

$$N_L\Delta\phi_2 = \int_0^{t_2} e_2\,dt \tag{1}$$

During the interval $0<t<t_2$, $e_2 = {}^1/_2\, v_s$. If the time phase angle at time t is defined as $\theta = wt$; then

$$N_L\Delta\phi_2 = \frac{1}{2w}\int_0^{\theta_2} |V_S|\sqrt{2}\sin\theta\, d\theta \tag{2}$$

To solve for θ_2, the angle at which core 2 fires

$$N_L\Delta\phi_2 = \frac{|V_S|}{\sqrt{2}w}(1-\cos\theta_2) \tag{3}$$

and

$$\cos\theta_2 = 1 - \frac{\sqrt{2}wN_L\Delta\phi_2}{|V_S|} \tag{4}$$

Determination of Second Firing Angle

In determining the second firing angle, it is convenient to define the difference between the flux levels existing in the two cores during the interval $0<t<t_2$ as $\Delta\phi_d$. This difference can be expressed in terms of the quiescent firing angle θ_q corresponding to zero control signal, and the angle θ_2. To advance the firing of core 2 from θ_q to θ_2, its flux level must be increased by ${}^1/_2\Delta\phi_d$ since the control signal produces equal but opposite flux change in the other core. Thus, following the same method used to obtain equation 2

$$\frac{1}{2}N_L\Delta\phi_d = \frac{1}{2w}\int_{\theta_2}^{\theta_q} |V_S|\sqrt{2}\sin\theta\, d\theta \tag{5}$$

and

$$N_L\Delta\phi_d = \frac{|V_S|}{w}\sqrt{2}(\cos\theta_2 - \cos\theta_q) \tag{6}$$

Equation 6 specifies the change in flux linkage which can take place in one load winding of core 1 between the angle θ_2, at which core 2 saturates, and the angle θ_1, at which core 1 saturates. Equating the integral of e_1 during the interval $\theta_2<\theta<\theta_1$ to this change in flux linkage.

$$V_S|\sqrt{2}(\cos\theta_2 - \cos\theta_q) = \int_{\theta_2}^{\theta_1} e_1\, d\theta \tag{7}$$

Thus the second firing angle can be obtained by solving equation 7 for θ_1.

The voltage e_1 is that which is induced in one load winding on core 1, and its mode of variation during the interval $\theta_2<\theta<\theta_1$ can be determined from the circuit shown in Fig. 8.

The circuit of Fig. 8 neglects the combined effect of forward resistance in the load rectifiers and resistance in the load windings of the saturated core (core 2). In this case $e_1 = v_{L1}$. Suitable resistors can be added to the circuit of Fig. 8 to represent the forward resistance of the rectifiers and the saturated resistance of the windings of core 2.[3]

Analogue Computation of Second Firing Angle

Manual calculation of θ_1 from equation 7 is tedious for any but the simplest types of load, whereas the use of an a-c network calculator greatly facilitates the plotting of many sets of characteristics showing θ_1 as a function of θ_2 for a wide range of circuit parameters and values of θ_q. A method for using a network calculator to solve for θ_1 is given in Appendix II.

Curves so obtained to relate θ_1 to θ_2 for various values of load shunt capacitance with a typical load are shown in Fig. 9. Note the action of the capacitance in delaying the firing angle of core

Vol.
Pt. 1
1956

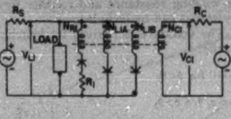


Fig. 8. Magnetic amplifier circuit when core 2 is saturated ($\theta>\theta_2$). Winding resistance and forward resistance of load rectifiers considered negligible

1. R_S was held at 100-ohm resistance throughout the analogue computations. The frequency of all voltages was 440-cps, hence the whole circuit was similar to that actually used with the magnetic amplifier of Appendix I.

While the method is inherently accurate, extreme accuracy was not attempted in obtaining these curves because of the lack of an accurate synchronous switch. Indicated values of θ_1 are probably within 4 degrees of their true value in most cases. The results presented here are in the nature of an illustration of the method of analysis and a qualitative comparison of the effects of various changes in the load impedance.

Manual Calculation of Second Firing Angle

Test results, Figs. 4(A) through 4(C), show that with the resistance load used, the second firing angle θ_1 changes only slightly from its quiescent value as the control signal is increased. This observation led to an investigation of the effect of load resistance on firing angles. The firing angle of the late-firing core 1, is related to the proportion of supply voltage applied to the load winding of core 1 after core 2 fires.

If K is the constant factor by which e_1 is increased after θ_2, then

$$e_1=\frac{K\sqrt{2}|V_S|}{2}\sin\theta \text{ for } \theta_2<\theta<\theta_1 \qquad (8)$$

The limits of K are $0<K<2$ as R_L varies from $0<R_L<\infty$. When $R_L=R_S$, then $K=1$ since both before and after the firing of core 2, core 1 will have the same voltage impressed.

From equation 7

$$|V_S|\sqrt{2}(\cos\theta_2-\cos\theta_q)=\int_{\theta_2}^{\theta_1}\frac{K\sqrt{2}|V_S|}{2}\sin\theta\, d\theta \qquad (9)$$

This equation agrees with Woodson's[3] equation F5. Solving now for θ_1

$$(\cos\theta_2-\cos\theta_q)=\frac{K}{2}(\cos\theta_2-\cos\theta_1) \qquad (10)$$

and

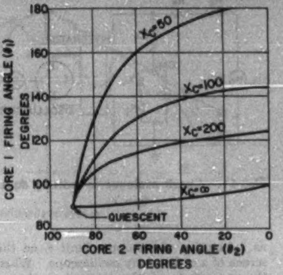


Fig. 9. Relationships between θ_1 and θ_2 for various values of X_C as determined from analogue computations ($R_S=100$, $R_L=150$, $X_L=150$), $\theta_q=90$ degrees

$$\cos\theta_1=\left(1-\frac{2}{K}\right)\cos\theta_2+\frac{2}{K}\cos\theta_q \qquad (11)$$

When K is less than unity, the equation is meaningful only for the values of θ_2 larger than that corresponding to $\theta_1=\pi$. For smaller values of θ_2, core 1 does not saturate, and a second firing angle θ_1 does not exist.

Equation 11 may be used to obtain the curves shown in Fig. 10 for various values of K. These curves show that as the load resistance becomes greater than the source resistance, K becoming larger, the control voltage loses control of the firing angle of the late-firing core.

Conclusions

The use of alternating voltage to control the half-wave bridge magnetic amplifier requires special precautions to avoid reduced output as a result of large control signals. The reduced output is caused by excessive voltage in the windings of the late-firing core during the set half-cycle.

This excessive voltage produces earlier firing of the core. Normally, the voltage impressed on the load windings of the late-firing core determines flux change on the set half-cycle and consequently the firing of the core. With large control signals, the control-winding voltage determines the flux change in the core as evidenced by a back voltage on two of the load winding rectifiers during the set half-cycle. The greater flux change as a result of the larger control signal then causes earlier firing of the core with a resulting reduction in output.

Limiting in the preamplifier is the most immediate solution to the problem of output reduction, although compromise design of the reset circuit holds further possibilities.

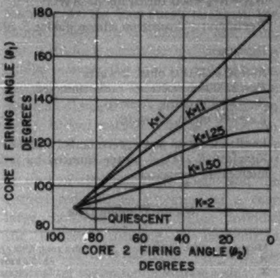


Fig. 10. Relationships between θ_1 and θ_2 for resistive load as calculated from equation 11

Test results with a resistance load show that under certain conditions of load the control voltage cannot control the firing angle of the late-firing core. Analysis shows that when the source and load resistance are equal, the control voltage has complete control of the firing angles of both cores. When the load resistance becomes greater than the source resistance the control voltage begins to lose control over the firing angle of the late-firing core. As the load resistance becomes much greater than the source resistance, the firing angle of the late-firing core tends to stay at its quiescent value even with increased control voltage.

The a-c network calculator can be used to determine the relationship between firing angles for any desired load condition. It is expected that the network calculator will find extended use not only in determining firing angles, but also in the choice of circuit elements by providing a more direct indication of the fundamental component of output voltage.

Appendix I. Design Details for Test Magnetic Amplifier

For the circuit of Fig. 1, the following listed components were used:

Excitation (V_S): 115 volts rms, 400 cps.

Core: Arnold Engineering Company *4635-D2* Deltamax toroid.

Core dimensions in inches:

Outside diameter, 1.375.
Inside diameter, 1.00.
Width, 0.25.
Tape thickness, 0.002.

Windings:

$N_{L1A}=N_{L1B}=N_{L2A}=N_{L2B}=1{,}300$ turns, no. 32 AWG (American Wire Gauge) wire
$N_{R1}=N_{R2}=500$ turns, no. 33 AWG

$N_{C1} = N_{C2} = 400$ turns, no. 33 AWG

Rectifiers: Transitron silicon diodes *S-201*.

Resistors:

$R_1 = R_2 = 27{,}000$ ohm, 2 watts
$R_3 = 2{,}000$-ohm 2-watt potentiometer
$R_C = 10{,}000$ ohms, 2 watts
$R_S = 100$ ohms, 10 watts

Load: Control phase, terminals 1 and 3 of a Kearfott *R-118-1A* motor shunted by a 2-microfarad capacitor.

Motor Winding Data

Maximum rated voltage	26 volts rms, 400 cps
R	98.5 ohms
X	123 ohms
Z	157 ohms
Power factor	0.63

Appendix II. Method of Using A-C Network Calculator in Determining Second Firing Angle

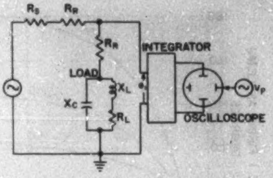


Fig. 11. Circuit analogue for solving equation 7

The upper limit of the integral given in equation 7 can be determined for given values of θ_q and θ_2 by displaying this integral as a function of its upper limit θ_1 on the screen of a cathode-ray oscilloscope. When the magnitude of the integral reaches a value predetermined in equation 7 for given values of θ_q and θ_2, the particular value of θ_1 corresponding to the given data is read from the oscilloscope.

The circuitry used in developing the required voltage integral, and a null method of reading the time phase angle θ_1 is illustrated in Fig. 11, which is a revision of Fig. 8 showing a switched alternating voltage applied to a circuit containing the R_S and load elements. In addition, two elements R_R have been added. If these are each made equal to the sum of the resistance of one "saturated" winding and the forward resistance of one load rectifier, the effects of winding resistance and rectifier forward resistance can be included.

The voltage v_t is derived by switching v_s so that

$v_t = 0$ for $\theta < \theta_2$

$v_t = v_s$ for $\theta > \theta_2$

The integrator and the vertical deflection system of the oscilloscope are calibrated by using v_s in place of e_1 at the integrator input and observing the peak to peak deflection of the oscilloscope trace. Phase calibrations are made with the use of a separate generator whose output voltage v_p can be phase-shifted relative to v_s by means of a calibrated phase dial. All phase angles can then be read on the phase dial by using it to move the oscilloscope trace horizontally to a predetermined null position.

References

1. An Improved Magnetic Servo Amplifier, C. W. Lufcy, A. E. Schmid, P. W. Barnhart. *AIEE Transactions*, vol. 71, pt. I, 1952, pp. 281–89.

2. Design Considerations of the Half-Wave Bridge Magnetic Amplifier, C. W. Lufcy, H. H. Woodson. *Ibid.*, vol. 73, pt. I, July 1954, pp. 220–26.

3. Preliminary Analysis of the Half-Wave Bridge Magnetic Amplifier, H. H. Woodson. *Navord Report 3596*, U.S. Naval Ordnance Laboratory, White Oak, Md., Feb. 1954.

Microwave Facilities with Built-In Reliability

R. G. KUCK
NONMEMBER AIEE

THIS PAPER describes a short-haul microwave radio system now providing toll quality telephone message service between San Diego (elevation 320 feet), Calif., and the Laguna Mountain recreation area (elevation 6100 feet), Calif. The microwave portion of the system employs Motorola frequency diversity equipment operating in the 5,925 to 6,425 mc (megacycle) common carrier allocation. The microwave system is presently multiplexed with 28 channels of Lenkurt *45 BX* carrier telephone channelizing equipment.

This system has been engineered so that it can be extended as required. Such an extension to the community of Borrego in the Imperial Valley is scheduled for completion in the second half of this year. At that time, a community dial office will be served exclusively by microwave facilities; Fig. 1.

The flexibility provided by such a combination of microwave and carrier equipments can be readily observed in that the original system will carry both the 28 channels to Mt. Laguna, and also the additional requirement of 12 channels to Borrego.

Initial Planning

The large increase in requirements between San Diego and Laguna Mountain, together with the lack of an existing readily expandable facility, opened the way for a primary relief study. Various combinations of cable and open-wire carrier systems as well as microwave radio were considered. The existing route consisted mainly of an open-wire lead which past experience had shown to be vulnerable to the high winds and sleet conditions in the Laguna Mountain area. A cable route along the existing wire lead would no doubt provide increased reliability as compared to the existing wire. However, in order to provide for this increased reliability, considerable pole line rearrangements would be necessary. Such rearrangements are entirely practical, but the charges for construction in the rugged mountain country would be high and no real diversification of routing would result.

The air-line distance between San Diego and Laguna Mountain is 44 miles, and a profile of this path is shown in Fig. 2. As may be seen, the controlling intermediate terrain features are located 31 and 41 miles respectively east of San Diego. Under varying conditions of the refractive index of the atmosphere, either of these points could present a serious obstacle to microwave propagation. Further field studies were undertaken to verify the information taken from the U. S. Geological Service maps. Following the map study height-loss testing equipment was employed to assist in determining the clearance required above the intermediate terrain features which might adversely affect microwave propagation on this path. By conducting height-loss

---

Paper 56-659, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 26, 1956; made available for printing April 27, 1956.

R. G. Kuck is with The Pacific Telephone and Telegraph Company, Los Angeles, Calif.

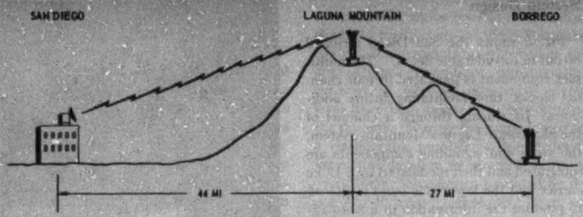


Fig. 1. Pictorial of microwave system

test transmissions, it was possible to verify data previously obtained from line-of-sight tests using the usual light-flashing techniques. From these field and office studies a tower height of 108 feet was selected. With this tower height the center of the reflector is 104 feet above mean ground level.

It might be surprising to discover that a site location in southern California would be subject to extreme weather conditions. When the moisture-laden air flowing across the San Diego lowlands rises to the cooler mountain areas, high winds and serious icing conditions often result. Recognizing these extreme weather conditions led to designing the tower, guys, anchors, and reflector as shown in Fig. 3, to withstand actual wind loads in excess of 100 miles per hour, and one inch of radial ice. The tower design also contemplated the additional load of a second reflector which would facilitate extension of the system to other locations as future requirements should dictate.

Terminal Design

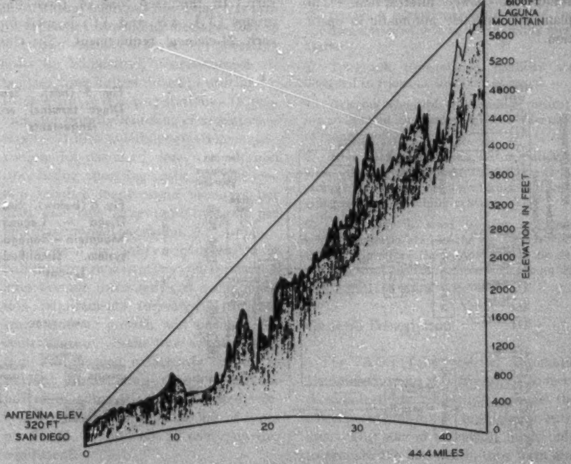
The San Diego terminal consists of 28 channels of Lenkurt *45 BX* carrier telephone equipment applied to the base band of the Motorola microwave equipment. At this terminal, remote type of microwave equipment was selected, thereby allowing for the installation of the carrier and microwave remote control bays on a lower floor of the office, with the installation of the radio-frequency and intermediate frequency portions of the microwave in a weatherproof cabinet on the roof of the building; see Fig. 4. All necessary tuning adjustments can be remotely actuated from the control unit located within the office. In this manner, wave guide runs are reduced to a minimum.

In consideration of the path length involved, the low power output (0.1 watt), and the possibility of severe fading as experienced on other microwave paths in this locale, three precautionary measures were employed to insure the desired reliability. This installation as previously noted is arranged for frequency diversity operation, i.e., two transmitters operating on different frequencies are simultaneously energized. The advantage gained by this technique arises from the small time differential normally existing between the times when a given fade crosses two adequately separated frequencies in a band. By means of proper control circuits at the receiving terminal, the base band signal as derived, can be taken from either receiver, depending upon the instantaneous received signal level. An automatic selection circuit using the information gained from the intermediate frequency limiters, together with a fail-safe trigger derived from a pilot receiver, accomplishes the required switching.

The second precautionary measure employed involves the use of high-gain antennas. At the San Diego terminal, the choice, based on availability and cost, was a 10-foot parabolic antenna. In a similar manner, the antenna design chosen for the mountain location consists of a 40-inch diameter parabola feeding an 8- by 12-foot curved reflector. The operation of this system may be better understood if the parabola is thought of as part of the feed mechanism, and the curved reflector as a section of a very large parabola. The reflector provides the ultimate focusing of transmitted energy and it is therefore the principle control on the antenna system gain.

The third means of providing system margin consisted of the installation of compandors on all voice circuits. The Lenkurt *45 BX* equipment has a compandor option and the additional cost associated with this option was thought to be justified for this application. Although, no appreciable increase in fading margin results from the use of compandors, a very noticeable decrease in the channel noise level for a given fade depth does result. This decrease in channel noise exists only when the received signal (prior to detection) is considerably above noise level. When this signal-to-noise ratio of a frequency modulated system approaches low values, the system rapidly degrades, and the compandor no longer provides its normally available quieting effect. Compandors, however, by allowing noise to come through only in the presence of speech and not during inter-

Fig. 2. San Diego–Laguna Mountain profile



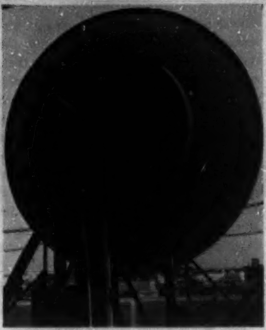


Fig. 3. San Diego, equipment cabinet and antenna

vals between syllables and words, reduce the disturbing effects on a voice channel when high background noise levels exist.

With one exception, the terminal installation on Laguna Mountain is similar to that at San Diego, already discussed. The radio equipment is of the micropackage variety, i.e., it is completely self-contained with the parabolic antenna located atop the equipment bay and pointed through a skylight at a reflector. In this instance, the skylight is specially designed to pass microwave energy with negligible impairment. By this means, the distance between the radio-frequency bay and the parabola feed is held to one foot or so. At this location, an emergency gasoline powered electric plant was installed to bridge the period of commercial power interruption. This plant is completely automatic in operation.

System Design

Fig. 5 shows the San Diego terminal layout in considerable detail. The multiplex equipment is packaged in four channel blocks, thus facilitating future additions. In tracing through a channel of the San Diego–Laguna Mountain system, the voice and signalling components are integrated and then modulated by a 12-kc carrier and the lower side band is filtered to produce the information in a 8- to 12-kc bandwidth. In similar manner, channels 2, 3, and 4 are modulated around 16, 20, and 24 kc respectively, to produce a total of four channels in the frequency spectrum of 8 to 24 kc.

In the pregroup unit, the first four channels are combined and modulated with a 96-kc carrier to produce a lower side band extending from 72 to 88 kc. In a similar manner, two other pregroup units employing modulating frequencies of 80 and 64 kc respectively produce lower side bands extending from 40 to 72 kc. At this point, 12 circuits have been modulated into the 40- to 88-kc range.

The informational energy now contained in the 40- to 88-kc band is sent through a base group unit where depending upon whether group 1 or 2 equipments are provided, a 40- to 88-kc or a 92- to 140-kc band containing 12 channels is generated. When the full complement of channel and pregroup units has been employed, a complete 24-channel group results, occupying a frequency spectrum from 40 to 140 kc.

Where the circuit requirement to be satisfied is 72 channels or less, only one additional step of modulation is necessary. In this step, one of three line groups (*XA*, *XB*, and *XC*) is used for each 24-channel requirement. By this means the 24 channel groups are stacked in one of the following allocations: *XA*, 40 to 140 kc; *XB*, 164 to 264 kc; and *XC*, 344 to 444 kc.

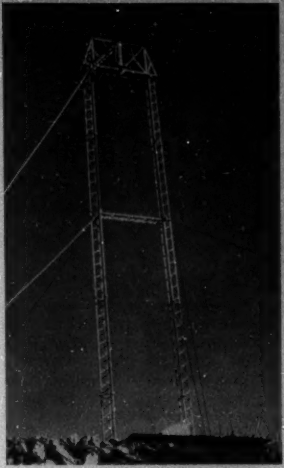


Fig. 4. Laguna Mountain, tower and reflector mounting

The total of the allocations shown is coupled into the base band or video amplifiers of the two Motorola diversity transmitters. The output of each base band amplifier, in turn, modulates the repeller of the klystron microwave output stage. The radio frequency wave thus generated is composed of a frequency-modulated carrier operating in the 6,000 mc range.

At the receiving terminal, the incoming 6,000-mc carriers are demodulated to 70 mc in two receivers and then in turn demodulated to the video or base band frequencies. By means of a comparator circuit the outputs of the receivers are compared on the basis of the better operational signal and, by means of series of coaxial relays, the desirable one is selected. By a process which is the reverse of that previously discussed, the individual channels are derived.

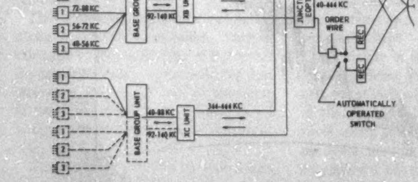
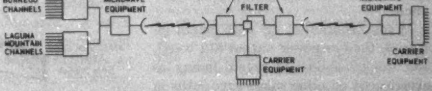

PRE-GROUP UNITS
RESERVED FOR BORREGO
COMPANDOR
CHANNEL UNIT
VF
S&G
1
2
3
1
2
3
72-88 KC
56-72 KC
40-56 KC
72-88 KC
56-72 KC
40-56 KC
BASE GROUP UNIT
BASE GROUP UNIT
BASE GROUP UNIT
40-88 KC
92-140 KC
XA UNIT
XB UNIT
XC UNIT
40-140 KC
164-264 KC
344-444 KC
MOTOROLA MICROWAVE EQUIPMENT
ORDER WIRE
ORDER WIRE
JUNCTION EQPT
40-444 KC
TRANS
TRANS
REC
REC
10' PARABOLA
AUTOMATICALLY OPERATED SWITCH

Fig. 5 (left). San Diego terminal arrangements

Fig. 6 (below). San Diego – Laguna Mountain – Borrego system. Simplified layout


SAN DIEGO
LAGUNA MOUNTAIN
BORREGO
CARRIER EQUIPMENT
BORREGO CHANNELS
LAGUNA MOUNTAIN CHANNELS
MICROWAVE EQUIPMENT
MICROWAVE EQUIPMENT
FILTER
CARRIER EQUIPMENT
MICROWAVE EQUIPMENT
CARRIER EQUIPMENT

Operation

Fig. 6 illustrates a simplified equipment layout from the San Diego terminal through the Mt. Laguna repeater to the Borrego terminal. A total of 40 circuits is now required, 28 terminating at Laguna Mountain and 12 extended to Borrego. The carrier equipment provided readily allows for augmentation to care for requirements which may be anticipated in the future.

The first section of the system was placed in service in March 1955. Initial system performance tests showed that all channels easily met the Bell Telephone System requirements for short-haul toll applications. In addition, continuous recordings of received microwave carrier level have been taken from last April to the present. During the summer months a considerable number of deep fades of very short duration were recorded. However, due to the margins provided by the use of diversity techniques, large antennas, and compandors as previously discussed, no noticeable system degradation was observed during this period.

Conclusions

Short-haul microwave systems (as contrasted with large-capacity transcontinental equipments) provide the design engineer with another very useful communications tool. In most instances, diversification of facility and routing, flexibility, and provision for future growth, together with immediate cost economies result from the use of microwave. Such systems properly designed and installed provide reliability comparable with other types of communication facilities.

Automatic Dispatch System for Half-Duplex Teletypewriter Lines

J. L. MAXWELL
ASSOCIATE MEMBER AIEE

H. K. FARRAR
MEMBER AIEE

HALF-DUPLEX multistation teletypewriter lines efficiently meet the needs of several points that require record communication with each other. In such cases it is important to be able to send messages to selected stations only. This ability is necessary to save paper, to provide privacy, and to avoid the need to scan all messages at each station to find those addressed to it. Direct interstation communication avoids sending all messages to a relay center for retransmission to other stations. This saves line and attendant time. To facilitate sending messages to addressee stations only, manually operated selective calling devices have been used. By present standards these are relatively slow. In addition, until recently it was necessary for attendants to initiate transmission of messages only when the line was idle. On a busy line this frequently required waiting for the line to become idle as shown by a line lamp. The attendant could then seize the line, operate the selecting equipment, and send the message. Sometimes attendants at two or more stations attempted to seize the line simultaneously. Unsuccessful attendants, unfortunately, had to resume waiting. On a busy line this procedure required appreciable amounts of nonproductive line and attendant time. Furthermore, there was no way to give an urgent message priority over other messages.

Automatic operation as described here greatly increases efficiency of half-duplex multistation teletypewriter lines. It provides privacy by sending messages only to addressee stations. It reduces operating costs by eliminating nonproductive attendant time hitherto used in waiting for the line and in selecting stations. It provides for priority handling of urgent messages. It also avoids contention among stations for use of the line. In addition to reducing operating costs, the new system handles many more messages per hour and can serve more stations per line than manual operation.

The need for automatic operation of half-duplex multistation teletypewriter lines became more acute as communication requirements increased with general economic growth and business decentralization. Some time ago the situation was deemed sufficiently urgent to warrant undertaking design work for providing automatic medium-sized teletypewriter services with the use of types no. *14*, no. *15*, and no. *19* teletypewriter equipment.

Basic Objectives

The following objectives were to be met.

The system should be automatic in:

1. Allowing all stations access to the line in specified sequence.
2. Permitting heavy-usage stations more frequent access to the circuit.
3. Directing messages only to addressee stations.

Centralized control should be provided.

Selection of stations should be by means of the standard teletypewriter 5-unit code for the following reasons:

1. Messages could be directed by station codes perforated in message tapes. This would permit use of the teletypewriter transmitter for sending station-selecting codes as well as messages.
2. Economies could be obtained by using the receiving teletypewriter as part of the selecting device as well as for receiving messages.
3. Telegraph transmission facilities are designed to transmit this type of signal.
4. Existing telegraph transmission measuring equipment and techniques could be used in maintenance.
5. It is adaptable to automatic switching with the reperforator-transmitter as a storage device between incoming and outgoing lines at switching centers.
6. This code is used by the sequence selector incorporated in the no. *28* teletypewriter; therefore the equipment to be developed would be compatible with the no. *28* line of teletypewriter equipment.

General Description

The TADS (Teletypewriter Automatic Dispatch System) was developed to meet these objectives. It greatly increases the efficiency of half-duplex multistation services. The station attendant needs only to prepare the message in tape form and

---

Paper 56-646, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 26, 1956; made available for printing April 17, 1956.

J. L. MAXWELL and H. K. FARRAR are with The Pacific Telephone & Telegraph Company, San Francisco, Calif.

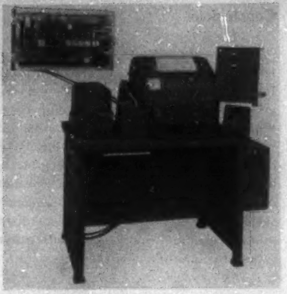


Fig. 1. Master station, controller cover removed

insert it in the transmitter of the teletypewriter, whether or not the line is busy. The TADS equipment starts the transmitters in sequence and directs the messages to the station or stations whose codes are perforated in the tapes or whose selecting buttons are depressed.

The basic TAD system consists of an automatic transmitter-start unit or controller located at a master station and a teletypewriter selector associated with each teletypewriter. The teletypewriters at all sending locations are equipped with perforators for tape preparation and with transmitters for automatic tape transmission.

Fig. 1 shows the master station teletypewriter with the station selector mounted on the right end of the table and the controller in a separate cabinet.

Twenty letters of the alphabet are suitable for station codes. V, M, O, and T are excluded as they can be generated by accidental opening and closing of the line. H and S are used for other functions. The same station codes are used to select stations and to start transmitters. The codes are used in lower case for station selection and in upper case for transmitter-start.

When 2-digit station codes are used, only two letters can be used as first digits on any one line. For example, X and Y might be used as first digits. In this case one group of 18 stations would have X for the first digit and a second group of 18 would have Y. This permits assigning 36 stations on one line.

The flexibility of the TAD system permits variations in arrangements to meet widely different customer requirements. This paper describes only the basic features.

Operation

Station codes are sent to all stations in sequence to start their transmitters if tapes have been placed in them. Stations which are not waiting to transmit respond to start codes with a no-tape signal (letter O). Thus each station in operating condition responds, either with transmission where tape is available or with a no-tape signal where there is no tape in the transmitter. This provides circuit assurance.

The search for waiting tapes proceeds as follows: Following a momentary idle line condition, the controller opens the line for about one second. This connects all machines to the line and starts their motors. The controller then sends the figures-shift code shifting all machines to upper case. This closes a set of contacts preparing each selector to respond to its transmitter-start code. The controller then sends the code for the first station in the sequence. If tape is available at this station, the transmitter starts. If no tape is available, the station sends the no-tape signal to indicate that the line and station are in working order. The controller then sends the start code for the next station. Should a station fail to respond with a no-tape signal or a message, the controller sends the start code a second time and, failing to get a response, he operates visual as well as audible alarms at the master station.

The operation of a SKIP key at the master station will retire the alarm, arrange to intercept messages for the station failing to respond, and cause the search to be resumed. Messages for that station will now be received on an intercept typing reperforator. The controller will continue to send the start code for this station in sequence but will not wait for a response before passing to the next. Continuing to send the code for a station on SKIP permits that station to send a message to notify the master station when service is restored. The SKIP key can then be restored to normal. The intercept arrangement will be described later.

Any combination of stations may be selected by including their codes in the tape. Each station can be equipped for one group or broadcast code. Any number of messages may be included in a single continuous tape, provided the proper format is used for each message. The transmitter will stop at the end of each message. The controller then sends the next start code. The transmitter stopped with messages remaining in its tape will restart and send the next message when it receives its start code in the next search cycle. This prevents one station from using the line for protracted periods. The equipment can be arranged to start designated stations more often than other stations if it is so desired.

The system has a normal capacity of 20 stations per line using single-digit codes, and a capacity of 36 stations per line using 2-digit codes. Transmission is limited to 60 or 75 words per minute at this time.

Station Selector

The station selector is housed in a cabinet mounted on the right end of the teletypewriter table; see Fig. 1. The equipment consists of relays, keys, lamps, and a cold-cathode-tube timing circuit. It is mounted on a gate which may be swung out to give access to the rear for maintenance or removed to facil-

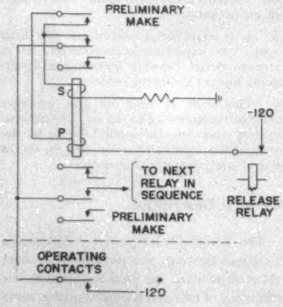


Fig. 2. Two-step relay circuit

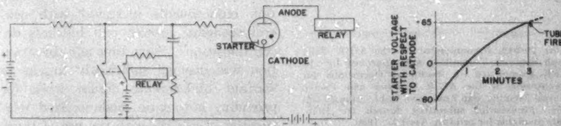


Fig. 3. Cold-cathode-tube timing circuit using battery reversal on capacitor

itate wiring. Components performing more than one function and special mounting arrangements have been utilized to reduce space requirements. D-c 120-volt power is obtained from the teletypewriter rectifier. The demountable line relay in the teletypewriter serves both the selector and the line magnets of the machine. The priority bid, circuit-assurance, and break signals for starting stations under manual operation are all generated by the same signal generator using different voltage, resistance, and capacitance values with a common polar relay.

Two-step relays have been used in some cases to make one relay do the work of two. An example of this technique is shown in Fig. 2.

Varistors are used in the circuit for two purposes:

1. To isolate operating paths, particularly those connected in multiple from a single contact in the typing unit.
2. To increase the release time of relays without affecting their operation. In this application the varistor is connected across the relay winding to present a high resistance to the operate potential and a low resistance to the voltage generated by the collapsing field.

Six special contact assemblies have been added to the typing unit to obtain the necessary functions for selective calling. A 3-element cold-cathode-tube timing circuit is provided to turn the station off one second after end of transmission. This operation may also be made dependent upon both end-of-transmission and end-of-message code. Making this operation dependent on end of transmission only has the advantage of automatically restoring the circuit assurance feature should the end-of-message code be omitted.

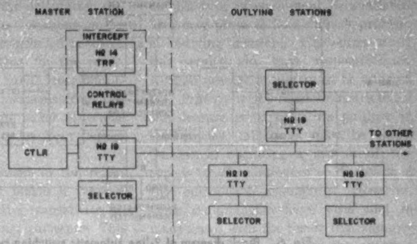
Three motor control options are available:

1. The motor turns on every time the station is connected to the line and turns off every time it is disconnected.
2. The motor normally stays on whenever the circuit is busy, regardless of whether or not the station has been selected. The motor automatically turns off whenever the circuit is idle for 15 seconds or more.
3. The motor runs continuously during service hours.

An audible alarm is activated and the transmitter stops when:

1. A tape snarl lifts the tape stop lever.
2. The end-of-address code is inserted in the tape ahead of the station selecting codes.
3. The code of the sending station has been omitted.

Fig. 4. Block diagram of TAD system with intercept equipment



After any of these events, the controller resumes searching after a momentary pause and searches the succeeding stations in sequence. Whenever the station where the fault has occurred is subsequently searched, it responds with a no-tape signal until such time as the tape is reset. The audible alarm can be retired only by removing the tape from the transmitter.

Should the controller be at rest after having made a complete search without finding a tape waiting, the busy lamp will be dark at all stations. During this condition, any transmitter can be started, without waiting for the automatic start, by operating the control key.

Such operation is recommended only in emergencies or in the event that the controller is not functioning. Otherwise, contention among stations may result and the advantage of automatic operation be lost.

Priority

Any station may send a priority message by the insertion of a tape in the transmitter and momentarily operating a key provided for this purpose. This causes a priority relay in the selector to operate and lock to the tape-stop contacts of the transmitter, which prepares the station to make a priority bid after the next end-of-message code.

Should the circuit be busy, the priority message will be the next transmitted. One nonpriority message may be transmitted ahead of the priority message if the controller is searching or the circuit is idle when a priority key is operated. In either case the following action takes place following the next end-of-message code:

1. The controller opens the line for the usual interval preceding the search. This causes the tape-withhold relays to be held operated through the marking contacts of the line relays at all stations. But at the station where the priority key was operated, a parallel holding circuit is provided for this relay through contacts of the priority relay.
2. The priority station (where the priority key was operated) opens the line a second time, releasing the tape-withhold relay at all other stations. This prepares the selectors at nonpriority stations to respond to start codes with no-tape signals even though tapes are available, which causes the nonpriority stations to be by-passed until the priority station has sent a message. Should the priority key be operated at more than one station the transmitter at the first such station in the search sequence will start. Following this transmission, the remaining such stations make a new priority bid without further attention. This arrangement permits separating the traffic into two classifications with very little loss of circuit time.

Controller

The controller is housed in an apparatus cabinet mounted on the wall or on a stand at the master station; Fig. 1. The components are mounted on a gate which swings out, giving access to the rear for maintenance. Components include relays, a 6-level-bank 22-terminal stepping switch, cold-cathode tubes, and associated resistors and capacitors. D-c 130-volt power is obtained from two

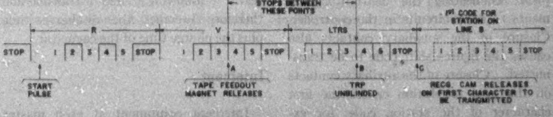

RECEIVING CAM STOPS BETWEEN THESE POINTS
R
V
LTRS
1ST CODE FOR STATION ON LINE B
STOP
1
2
3
4
5
STOP
A
B
C
START PULSE
TAPE FEEDOUT MAGNET RELEASES
TRP UNBLINDED
RECG. CAM RELEASES ON FIRST CHARACTER TO BE TRANSMITTED

Fig. 5. Typing reperforator control signal timing

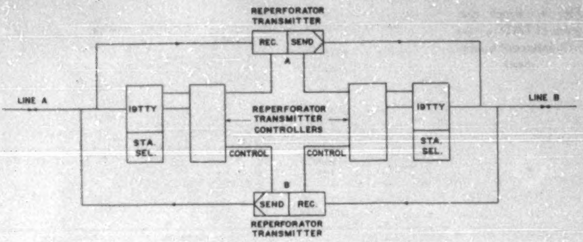


Fig. 6. Block diagram of 2-line automatic switching system

selenium rectifiers in the controller. Timing circuits, 2-step relays, and other devices similar to those in the selector are used in the controller.

The station codes are wired to five of the six stepping-switch levels for single-digit codes. In the case of 2-digit codes, the second digits only are wired to the bank.

The automatic transmitter-distributor of the teletypewriter at the master station serves to transmit both messages and start codes. This transmitter-distributor is equipped with auxiliary contacts operating from a cam on the lower end of the distributor shaft and with tape-withhold parts. Operation of the tape-withhold parts prevents stepping of tapes while start codes are being sent. Each time the brush arm of the distributor enters the start segment, the auxiliary contacts close to control the various stepping and transfer functions.

OPERATION

A relay at the master station follows all line signals. Upon closure of the line following transmission of a message this relay operates to its marking contact. This causes a positive charge to build up on the control anode of a cold-cathode tube in the associated timing circuit. The tube fires in about two seconds, opening the line. This releases a relay and lights the "busy" lamp at each station. It also starts a second timing circuit that after about one second closes the line, connects all machines to the line and starts the master station transmitter. The auxiliary contacts in the transmitter close at the beginning of the start pulse, connecting the distributor segments to relay circuits in the controller. Closure of the auxiliary contacts also causes the figures shift code to be transmitted. Closure of the auxiliary contacts on the next revolution causes the first character of the station code, for example, X, to be transmitted. This assumes the equipment is arranged for 2-digit calling and that X is the first digit of the first group. On the next revolution the distributor is connected to the stepping-switch bank and the second character in the station code is sent, for example, A, if station XA is to be started. In this case the 1 and 2 levels of the stepping-switch would be wired for marking and the 3, 4, and 5 levels for spacing. If a no-tape signal is returned indicating everything in order but no tape available, the X is again transmitted. The switch steps again, and the next station code, for example, B, is sent. Assuming tape is available, the transmitter starts and addressee stations are selected as a result of station codes in the tape. The controller now waits until end of transmission before sending the next station code.

At the end of one cycle, the stepping switch operates a relay which causes the first digit for the second group, for example, Y, to be transmitted at each step. Should a complete search be made of both the X and Y groups without starting a transmitter, the controller will shut down for a rest period of three minutes, six minutes, or for an indefinite period, as determined by a key at the master station. Should the controller be at rest it can be reactivated by operating a key at any station.

The cold-cathode-tube timing circuit includes an 18-microfarad capacitor and a 10-megohm resistor as shown in Fig. 3. To attain a 5-to-6 minute rest interval, the capacitor is charged in one direction while the line is busy and in the other during the rest interval. The timing interval following end-of-transmission therefore becomes the discharge time plus the charge time of the capacitor.

Intercept

Intercept equipment at each master station includes a typing reperforator, a station selector, control relays, and a panel equipped with a key and lamp for each station.

The typing reperforator receives messages with nonvalid station codes or incorrect format. Messages for stations temporarily out of service may also be intercepted.

A lamp lights on each step of the stepping switch in the controller indicating the progress of the search. Operation of the associated locking-type SKIP key causes the lamp to stay lit and the messages for that station to be received on the intercept typing reperforator. The start code for the station on SKIP is sent as before but the controller continues the search sequence even though no response is received.

Tapes received on the intercept reperforator can be retransmitted by means of the master station transmitter. Should incorrect station codes appear in the tape the correct codes can be perforated in a separate tape and the two spliced together. When push-button equipment is provided, it can be used to provide the correct codes without preparing a tape. Fig. 4 shows the TADS with intercept equipment.

Push-Button Selection

Push-button selecting equipment provides for sending station codes for any one or more of 20 stations, a reperforator for manual relaying, or a reperforator-transmitter for automatic retransmission to another line.

A key box equipped with key and lamp for each station, a common lamp, a start key, and a release key, is mounted on the left end of the table for a no. *19* teletypewriter. This adds about $2^1/_2$ inches to the over-all length of the machine.

Relay equipment for the push-button feature is housed in a 6-plate apparatus cabinet identical to that used for the controller. The basic push-button unit provides for eight codes. Provision is made for two additional units of seven codes each. The push-button equipment has the following limitations and operating features:

1. It is for use with single-digit station codes only.
2. There is a maximum of 20 station selecting codes.
3. It sends end-of-address codes after the station codes.
4. It sends end-of-message code if not in tape.
5. The transmitter can be loaded and station codes selected while the line is busy.

6. It permits proper direction of tapes with incorrect or garbled station codes. (The incorrect codes become part of the text.)

7. It permits sending tapes including proper station codes by simply operating the START key.

Push-button operation has advantages where:

1. Attendants cannot be depended upon to follow the prescribed message format.
2. Tapes without station codes are received over other facilities for introduction into a TAD system.
3. Messages with incorrect codes are received on an intercept reperforator and must be re-sent.

Typing Reperforator Control Circuit

This circuit facilitates the manual relaying of traffic between two circuits. Each circuit must be equipped at the relay point with a no. *19* teletypewriter, a selector, and a typing reperforator with its control circuit. The following requirement must be met:

No codes be punched in the tape that might interfere with directing the message on another circuit. The tape must be automatically fed out clear of the tape guide to permit it to be torn off without mutilating perforations needed for retransmission. The reperforator must resume operation without omitting a character, misprinting, or introducing an extraneous character in case it is selected for another message during feedout. This is accomplished by assigning a 2-digit code for the typing reperforator; for example RV. A letters-shift code following the letter V is inserted to allow time for the receiving cam of the reperforator to stop before a selecting or message character is received; see Fig. 5.

Automatic Cross-Office Switching Between Two Lines

Fig. 6 shows an arrangement for automatically relaying messages between two lines. Two reperforator-transmitters are required at the switching point, one for relaying in each direction. A teletypewriter and selector are also required here on each line. The switching point may be either at the master station or at an outlying station.

Messages are directed to the receiving side of a reperforator-transmitter by means of a station code. Receipt of the code connects the receiving side of the reperforator-transmitter to the line and disconnects the teletypewriter at that station. The message must, of course, include the codes for the destination stations on the other line. Stations on the originating line can be included provided their codes precede that for the reperforator-transmitter. Each reperforator-transmitter receives messages from one line and transmits them to the other line in response to start codes from the associated controller.

A flip-flop arrangement in the reperforator-control circuit makes it possible to have the reperforator-transmitter respond to alternate start patterns. This means that the transmitter can be started after each message or after alternate messages, whichever best meets traffic requirements. Preferential handling of priority messages is provided by use of a priority code that conditions the flip-flop circuit to start the reperforator-transmitter after each message. Should the outgoing line be idle, reception of a priority message by the reperforator-transmitter will cause a pulse to be sent to the line. This activates the controller which starts all machines and sends the figures-shift code causing the priority message to be transmitted.

Tape failure in the reperforator-transmitter will operate an audible alarm and automatically cause the teletypewriter to receive all cross-office messages.

Automatic cross-office relaying of messages extends the application of the TAD system making it feasible for as many as 72 stations (36 on each of two circuits) to exchange traffic automatically.

Conclusion

The widespread application of TADS equipment to half-duplex private-line teletypewriter services has emphasized the need for automation of moderately extensive services. Economies in attendance and paper usage and greatly increased circuit efficiency have resulted in demand for the system.

This type of service brings with it new maintenance problems for the serving communications company. Users frequently require stations at remote locations. This places more rigid requirements on line facilities which sometimes are open wire. Minor interruptions or irregularities which may pass unnoticed with manual operation may seriously interfere with an automatic system. An important requirement is that adequately trained maintenance people experienced on this type of equipment be available. This may prove to be a difficult problem in outlying districts. Customers using this service also have responsibilities. A rigid message format must be adhered to. Priority traffic must be restricted if this feature is to serve its purpose in emergencies. Messages must be short. Success in operating a system of this type is largely contingent on discipline of the customer's attendants. However, these requirements have been met in many cases with the result that the system described is meeting the exacting requirements of many customers.

Nomenclature

End-of-address code: A combination of characters or functions, normally carriage-return and line-feed, used between the station codes and the associated message text for disconnecting the nonselected stations.

End-of-message code: A combination of characters or functions, normally figures-shift and the letter H, used at the end of the message text to stop the sending transmitter and to cause the controller to resume the search.

Half-duplex line: A teletypewriter line capable of transmission in both directions but not simultaneously.

Mark: Closed circuit.

Master station: The station on a TAD system at which the automatic start unit (controller) is located. When intercept equipment is used it is also installed at the master station.

No. 19 teletypewriter: A page-copy teletypewriter, transmitter-distributor, and a tape perforator mounted on one table.

Perforator: A device for preparing a message in punched tape form for transmission by means of a transmitter-distributor (part of the no. 19 teletypewriter).

Reperforator-transmitter: A teletype machine consisting of a typing reperforator and automatic transmitter in one unit. Messages are received in the form of typed and punched tape and are automatically retransmitted. Separate control provides storage between the receiver and transmitter.

Space: Open circuit.

Transmitter-distributor: A device for automatic transmission of perforated tape on a teletypewriter circuit (part of the no. 19 teletypewriter).

Typing reperforator: A teletypewriter which receives messages from the line, types them on tape, and perforates the tape for retransmitting the messages.

The Analysis of Sampled Data Servomechanisms Performed on the IBM Type 650

B. M. TOSTANOSKI
ASSOCIATE MEMBER AIEE

Synopsis: Because of the need for more accurate and versatile control capabilities than can be obtained economically with continuous systems, discrete filters (digital computers) are being substituted as control elements in many applications. The exact response of linear systems which combine continuous and discrete filters can be obtained for discrete values of time by a series of well-defined mathematical processes. Since these processes require many computations of a repetitive nature to obtain complete response data, manual methods of computing are often not practicable and general-purpose digital computers such as the IBM (International Business Machines Corporation) type *650* are used to assist in the analysis of these systems.

A SAMPLED DATA control system is one in which the signal supplied to one or more parts of the system is delivered at discrete values of time. The types of sampled data systems which have attained importance in military and industrial control applications are those which use discrete filters (digital computers) as control elements to monitor the activity of continuous controlled members.

The response of a real linear system containing both continuous and discrete filters can be exactly defined at sampling instants by a difference equation having a finite number of terms. The constant coefficients for this difference equation are obtained directly from the discrete system transfer function. A discrete system transfer function is in turn obtained by mathematically substituting equivalent discrete filter transfer functions for continuous functions and then reducing the system representation to a ratio of polynomials in the discrete transform variable z.

System reduction and subsequent response evaluation require a large number of computations, even when simple systems are investigated. Therefore, general-purpose digital computers such as the IBM type *650* can be profitably applied to the solution of these problems.

The purpose of this paper is to discuss the development of computer programs which can be used to assist in obtaining fast and accurate solutions of linear sampled data systems. After the programs are discussed, some indication is given regarding their application to nonlinear sampled data systems and to continuous linear systems.

Filter Representation and System Reduction

In general, sampled data control systems combine both continuous and discrete filters. The linear continuous filter, Fig. 1(A), is defined by the differential equation

$$o(t)+B_1\frac{d}{dt}o(t)+B_2\frac{d^2}{dt^2}o(t)+\ldots+B_b\frac{d^b}{dt^b}o(t)$$
$$=A_0i(t)+A_1\frac{d}{dt}i(t)+A_2\frac{d^2}{dt^2}i(t)+\ldots+A_a\frac{d^a}{dt^a}i(t)\quad(1)$$

and by the transfer function

$$\frac{O(s)}{I(s)}=G(s)=\frac{\sum_{j=0}^{a}A_js^j}{1+\sum_{j=1}^{b}B_js^j}\quad(2)$$

The input to a continuous filter may be either a discrete or continuous function of time, whereas the output is always a continuous function.

The linear discrete filter, Fig. 1(B), is defined by the difference equation

$$o^*(t)+B_1o^*(t-T)+B_2o^*(t-2T)+\ldots+B_bo^*(t-bT)$$
$$=A_0i^*(t)+A_1i^*(t-T)+A_2i^*(t-2T)+\ldots+A_ai^*(t-aT)\quad(3)$$

and by the transfer function

$$\frac{O^*(s)}{I^*(s)}=W(\epsilon^{-sT})=\frac{\sum_{j=0}^{a}A_j\epsilon^{-jsT}}{1+\sum_{j=1}^{b}B_j\epsilon^{-jsT}}\quad(4)$$

where the asterisk is used to designate information which is in sampled form and T is the sampling period. Setting $\epsilon^{-sT}\equiv z$, equation 4 becomes

$$\frac{O^*(s)}{I^*(s)}=W(z)=\frac{\sum_{j=0}^{a}A_jz^j}{1+\sum_{j=1}^{b}B_jz^j}\quad(5)$$

The input and output of a discrete filter are discrete functions.

One of the simplest forms of a servomechanism which combines both types of filters is shown in Fig. 2(A). The continuous system input and the output of the controlled member are differenced to determine the system error. This error is sampled for processing within the discrete filter. The exact determination of the system response would require the inversion of transforms containing infinite series.[1] Truncating the series and then inverting yields a solution of the system response which is a continuous function of time. This method results in an approximate evaluation of the output for all values of time.

A more useful approximation of the system output can be obtained by mathematically reducing the entire system to an equivalent discrete filter described by a difference equation. This method results in an expression for the system output which is not continuous. However, the solution is exact at the sampling instants of time.

The mathematical technique by which discrete transfer functions are defined in terms of corresponding continuous functions[1] is applicable to the configuration shown in Fig. 3(A) when the input to the continuous filter is sampled data. The combination of a continuous filter followed by a sampler is reduced to an equivalent discrete filter, shown in Fig. 3(B).

For mathematical equivalence, if

$$G(s)=\frac{k_1}{s+a_1}+\frac{k_2}{s-a_2}+\ldots+\frac{k_n}{s-a_n}+\ldots+\quad(6)$$

then

$$\frac{O^*(s)}{I^*(s)}=G^*(s)=\frac{k_1}{1-\epsilon^{a_1T_s}}+\frac{k_2}{1-\epsilon^{a_2T_s}}+\ldots+\frac{k_n}{1-\epsilon^{a_nT_s}}+\ldots+\quad(7)$$

Summing the terms of equation 7 over a least common denominator will produce

---

Paper 56-680, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 22, 1956; made available for printing May 16, 1956.

B. M. TOSTANOSKI is with the International Business Machines Corporation, Endicott, N. Y.

The author wishes to express his appreciation to M. Kloomok, C. J. Hoppel, and M. M. Dickinson for their helpful suggestions and criticisms.

Vol. 7
Pt. 1
1956

a ratio of polynomials in s, having real coefficients, as in equation 5. When $G(s)$ contains multiple roots, the corresponding discrete equivalents can be described in an analogous manner.

By using the equivalence indicated by equations 6 and 7, a linear system can be totally represented in the domain of discrete filter operation. The over-all system transfer function can be reduced to the form of equation 5, which immediately defines the coefficients of the corresponding difference equation, equation 3. This method is apparent in the reduction procedure followed in Fig. 2.

The sampled output signal of Fig. 2(A) may be determined by first redrawing this figure as Fig. 2(B) in which the sequence of sampling and differencing is interchanged and the feedback path is redefined. The feedback path, having sampled input and output, is the equivalent of a discrete filter and can be represented as $G^*(s)$ by the method which resulted in equation 7. This reduction is incorporated in Fig. 2(C) which is obtained from the configuration of Fig. 2(B) by reducing the loop and adding a mathematical sampler after the remaining continuous filter. This sampler is placed on the output of the controlled member in order to permit the mathematical substitution of an equivalent discrete filter transfer function, $G^*(s)$, for the continuous $G(s)$. If this were not done the output of the system would be continuous and would have to be represented by an infinite number of terms, as previously mentioned.

Again applying the criteria for equivalence established by equations 6 and 7, the system shown in Fig. 2(D) is reduced to that of Fig. 2(E) from which the over-all discrete system transfer function can be defined as

$$\frac{C^*(s)}{R^*(s)}=\frac{D(z)G^*(s)}{1+D(z)G^*(s)} \tag{8}$$

By algebraic manipulation, equation 8 can be reduced to the form of equation 5, which for convenience is written as

$$\frac{C^*(s)}{R^*(s)}=\frac{\sum_{j=0}^{a}A_jz^j}{1+\sum_{j=1}^{b}B_jz^j} \tag{9}$$

Equation 9 is recognized as the transform of equation 3. The solution of equation 9 can be written immediately as

$$c^*(nT)=\sum_{j=0}^{a}A_jr^*[(n-j)T]-\sum_{j=1}^{b}B_jc^*[(n-j)T] \tag{10}$$

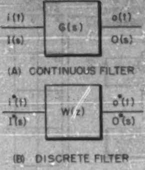


Fig. 1. Continuous and discrete-filters

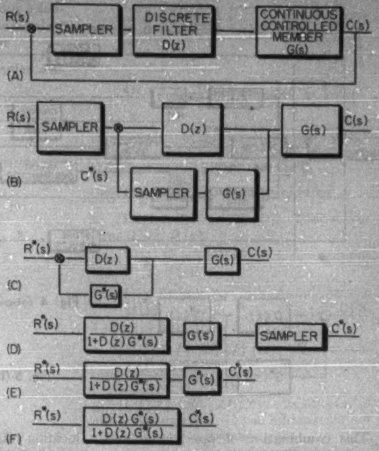


Fig. 2. System reduction

and the exact response can be evaluated at each sampling instant by $(a+b+1)$ multiplications and an equal number of additions.

A sampled response evaluated by this procedure must be computed at every sampling instant for monotonically increasing values of time. When it is necessary to evaluate the response at a given $t=nT$, the indicated operations must be performed $n+1$ times. If a and b are large, this presents a formidable manual-computing problem.

The problems of defining continuous filters as equivalent discrete filters, system reduction, and response evaluation are solved by elementary algebraic and logical processes which can readily be performed on a general-purpose digital computer.

System Reduction and Response Evaluation Performed on IBM Type 650

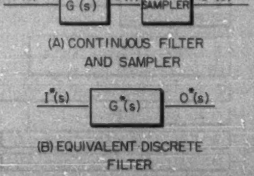
COMPUTER PROGRAMS

To minimize the time between definition and solution of sampled data systems a number of operational and logical programs are compiled to form a "library" routine, which is one designed for the solution of problems of a recurrent nature.

An operational program is one which will transform one set of data into another by arithmetic processes, as an example, one which will determine the real and complex roots of a polynomial of any order up to some specified limit. The only variations encountered in this problem are the polynomial coefficients. To obtain the roots of a given polynomial it is necessary only to enter the coefficients into predetermined storage locations and instruct the computer to begin the operations necessary to compute the corresponding roots. As the roots are obtained they are placed in sequential order in other preassigned storage locations.

The users of this program need have no detailed knowledge of the computer operations used to obtain the results. It is necessary only to know how to locate input data and where to find the output data. Results of a given input must either be displayed or moved to other storage locations before a new set of coefficients is processed.

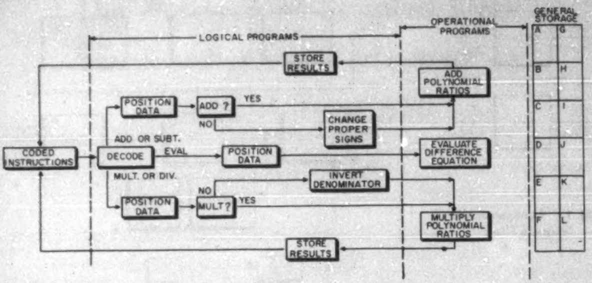
When a group of operational routines is collected or designed for the solution of a given type of problem it is profitable to develop logical programs to procure and position input data, relocate results, and select the sequence of operations. The logical program, in turn, can be controlled by coded instructions.



(A) CONTINUOUS FILTER AND SAMPLER

(B) EQUIVALENT DISCRETE FILTER

Fig. 3. Equivalent discrete filter definition


LOGICAL PROGRAMS
OPERATIONAL PROGRAMS
GENERAL STORAGE
STORE RESULTS
ADD POLYNOMIAL RATIOS
POSITION DATA
ADD ?
YES
NO
CHANGE PROPER SIGNS
ADD OR SUBT.
CODED INSTRUCTIONS
DECODE
EVAL
POSITION DATA
EVALUATE DIFFERENCE EQUATION
MULT. OR DIV.
INVERT DENOMINATOR
NO
POSITION DATA
MULT ?
YES
MULTIPLY POLYNOMIAL RATIOS
STORE RESULTS
A
G
B
H
C
I
D
J
E
K
F
L

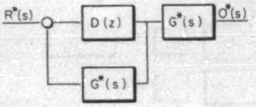

R*(s)
D (z)
G*(s)
O*(s)
G*(s)

Fig. 4 (above). Interpretive computer program

Fig. 5 (left). Sampled data system

This combination of operational and logical programs controlled by coded instruction is called an interpretive system and is designed to reduce the number of coded instructions required to solve a particular problem. One coded instruction may be used to control the selection and manipulation of large numbers of data.

When the logical program receives a coded instruction, it receives a series of numbers which it separates into groups, thus:

[*OP*] [*A*] [*B*] [*C*]

These numbers are decoded or interpreted for information which designates the operation to be performed, [*OP*], the general storage location of input data, [*A*] and [*B*], and the general storage location into which output data is to be placed, [*C*].

Once the operational and logical programs are compiled into a library routine, the only programming necessary is the coding of instructions applicable to a specific problem. The simplicity of writing these coded instructions will be illustrated in a later section in conjunction with Fig. 4 which outlines an interpretive system.

Equivalent Discrete Filters

The part of the library program used to reduce a continuous filter and sampler combination to an equivalent discrete filter is composed of programs designed to perform the following processes:

1. Determine the real and complex roots of a polynomial.
2. Determine the coefficients for the partial fractions[3] of equation 6.
3. Determine the coefficients of *z* in equation 7.
4. Add polynomial ratios over a least common denominator to alter the form of equation 7 to that of equation 5.

For a continuous transfer function having a fifth-order denominator this entire procedure requires only about 40 seconds of computing time on the IBM *650*.

Roof-finding procedures can be found in books[3] on numerical methods and the determination of partial fractions by computer methods can be found in reference 2. Determining the coefficients of *z* of equation 7 is a simple and straightforward process of function evaluation.

The addition of polynomial ratios

$$\frac{P_{11}(z)}{P_{21}(z)}+\frac{P_{12}(z)}{P_{22}(z)}=\frac{P_{13}(z)}{P_{23}(z)} \tag{11}$$

by interpretive methods is accomplished by an operational program designed to perform

$$\frac{P_{11}(z)P_{22}(z)+P_{21}(z)P_{12}(z)}{P_{21}(z)P_{22}(z)}=\frac{P_{13}(z)}{P_{23}(z)} \tag{12}$$

and a logical program which will obtain four sets of input data and store two sets of results if each polynomial is considered separately.

To keep the programming as simple and compact as possible, all polynomials are assumed to have *n* terms; zero coefficients are introduced to fulfill this assumption. The coefficients of the denominator are stored in order immediately after the coefficients of the numerator, and the interpretive program is designed to move groups of 2*n* units of data whenever an operation involves ratios of polynomials. The location of the first coefficient of a polynomial ratio is given in the coded instruction.

The operation code for the addition of polynomial ratios will be designated as [ADD *P*].

System Reduction

After all transfer functions are defined in the form of equation 5, the over-all system transfer function is obtained by two basic operations:

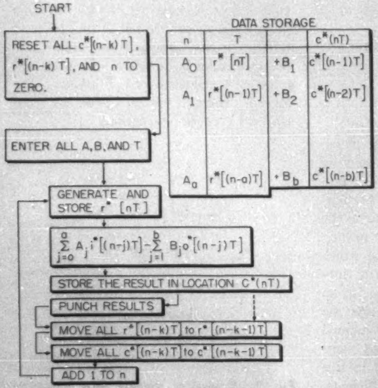

START
RESET ALL c*[(n-k)T], r*[(n-k)T], AND n TO ZERO.
ENTER ALL A, B, AND T
GENERATE AND STORE r* [nT]
$\sum_{j=0}^{a} A_j i^*[(n-j)T]-\sum_{j=1}^{b} B_j o^*[(n-j)T]$
STORE THE RESULT IN LOCATION c*(nT)
PUNCH RESULTS
MOVE ALL r*[(n-k)T] to r*[(n-k-1)T]
MOVE ALL c*[(n-k)T] to c*[(n-k-1)T]
ADD 1 TO n
DATA STORAGE
n
T
c*(nT)
A0
r*[nT]
+B1
c*[(n-1)T]
A1
r*[(n-1)T]
+B2
c*[(n-2)T]
Aa
r*[(n-a)T]
+Bb
c*[(n-b)T]

Fig. 6 (left). Program to evaluate difference equation

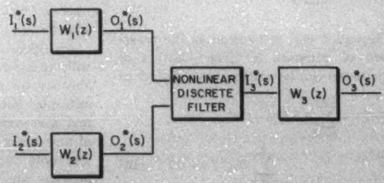

I1*(s)
W1(z)
O1*(s)
NONLINEAR DISCRETE FILTER
I3*(s)
W3 (z)
O3*(s)
I2*(s)
W2(z)
O2*(s)

Fig. 7 (right). Sampled data system containing a nonlinear element

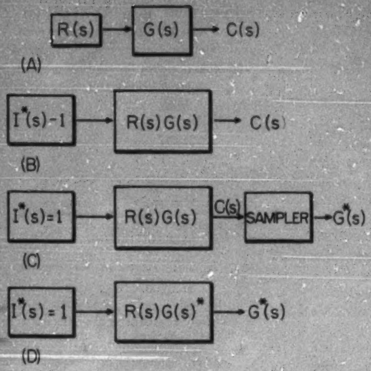
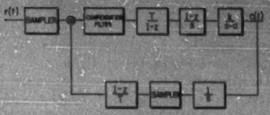


Fig. 8 (left). Reducing a continuous system to an equivalent discrete filter

Fig. 9 (right). Sampled data system



1. Adding polynomial ratios over a least common denominator.

2. Multiplying polynomial ratios.

For the system of Fig. 5 with $n = 10$, less than 10 seconds are required by the IBM *650* to determine the over-all system transfer function.

The multiplication of polynomial ratios [MULT *P*] is done by an operational routine designed to perform the operation

$$\frac{P_{11}(z)}{P_{21}(z)}\,\frac{P_{12}(z)}{P_{22}(z)}=\frac{P_{13}(z)}{P_{23}(z)} \tag{13}$$

The reduction of a feedback loop requires the following operation on a combination of polynomials

$$\frac{\dfrac{P_{11}(z)}{P_{21}(z)}}{1+\dfrac{P_{12}(z)}{P_{22}(z)}\,\dfrac{P_{13}(z)}{P_{23}(z)}}=\frac{P_{14}(z)}{P_{24}(z)} \tag{14}$$

By applying the program for the multiplication of polynomial ratios and then the program for addition, the denominator of the left-hand side is reduced to one ratio of polynomials. To complete the problem, a division of polynomial ratios must be performed. An operation code [DIV *P*] and its associated logic is designed to invert the denominator, then the same logical and operational routines which effect the multiplication of polynomial ratios may be used.

When positive feedback loops are encountered, it is necessary to subtract polynomial ratios. For this purpose an operation code [SUB *P*] and associated logic is designed to reverse the signs of the proper coefficients so that the logical and operational routines for addition of polynomial ratios may be used.

RESPONSE EVALUATION

If the transform of the sampled input can be specified in closed form, the system can be reduced by including the input transform with the system transfer function, thus defining an equivalent system whose input is a unit impulse at $t = 0+$. This simplifies the input function generation.

The response function, equation 10, is evaluated by the program shown in Fig. 6. When the system input is a unit impulse at $t = 0+$, the function generator enters a (+1) into the storage location assigned for $r^*(nT)$ at $t = 0$, and thereafter enters zero into this location. In general, the function generator must functionally or logically define $r^*(nT)$ for all values of n.

Since the function generator is part of the problem input data, this evaluation program will be complete, when the coefficients of the difference equation are properly placed. An operation code [EVAL] and its associated logic is established to move the coefficients of the transfer function into the evaluation program and begin evaluating the response at sampling instants.

The evaluation program is shown as a closed loop and, once entered, it continues to evaluate the function for successive values of time. When this loop characteristic is undesirable, simple program alterations are made to satisfy particular problem requirements.

To illustrate the simplicity of using interpretive computer programs, the system of Fig. 5 will be reduced and evaluated with five instructions whose operation codes involve polynomial ratio manipulation.

Assuming an interpretative routine as shown in Fig. 4, if

$D(z)$ is stored in general storage A,
$G^*(s)$ is stored in general storage B, and
(1/1) is stored in general storage C,

then the coded instructions would be

1. [MULT *P*] [*A*] [*B*] [*D*]. *D* now contains $G^*(s)\,D(z)$

2. [ADD *P*] [*C*] [*D*] [*E*]. *E* now contains $1+G^*(s)\,D(z)$

3. [DIV *P*] [*A*] [*E*] [*F*]. *F* now contains $\dfrac{D(z)}{1+G^*(s)\,D(z)}$

4. [MULT *P*] [*B*] [*F*] [*H*]. *H* now contains $\dfrac{D(z)\,G^*(s)}{1+G^*(s)\,D(z)}$

5. [EVAL] [*H*] [*X*] [*X*]. Evaluation of difference equation. [*X*] is information not applicable to this operation code.

NONLINEAR SYSTEMS

Theoretically, there is only one restriction placed on a sampled input; it must be finite in magnitude. There need be no correlation whatever between samples at $t = nT$ and $t = (n \pm 1)\,T$. Therefore sampled inputs to a linear discrete filter may be outputs of a nonlinear function generator.

By using the program in Fig. 6 as a unit building block to represent linear portions of a system, it is possible to simulate systems which contain nonlinear discrete elements with a minimum of effort, since all of the linear portions may be reduced and evaluated by the methods outlined. For example, to represent the system of Fig. 7, the only additional programming necessary would be to simulate the nonlinear discrete filter and to open the loop in the evaluation program.

CONTINUOUS SYSTEMS

Although the discussion to this point has been directed towards the solution of sampled data systems, the programs considered may be applied to continuous linear systems.

The programs concerned with operations on polynomial ratios are applicable for the reduction of continuous systems to a single ratio of polynomials which can be manipulated to reveal response data.

When applying classical methods to evaluate the response of a system defined by

$$C(s)=R(s)G(s)=\sum_{j=1}^{n}\frac{k_j}{s-a_j} \tag{15}$$

it is usually necessary to compute exponential and trigonometric functions for each time point.

By defining an equivalent $C^*(s)$ for $C(s)$, as shown in Fig. 8, trigonometric and exponential functions need be evaluated only once for a given system, system input, and desired sampling period to obtain the coefficients of z in equation 7. Then the solution of the system can be reduced to repetitive evaluation of equation 10. Evaluating the response of continuous systems by this method reduced computing time by a factor of about 10.

Since the difference equation representing the continuous time function is mathematically exact at sampling instants, the accuracy of the computed solution is affected only by performing arithmetic operations on numbers which have been rounded to 8 or 10 digits. The truncation error introduced by applying conventional numerical techniques to the solution of linear differential equations is eliminated.

Conclusion

The library programs discussed can be designed for the IBM *650* to obtain fast and accurate solutions of sampled data systems arising in the design of discrete control elements. One example of the systems studied is shown in Fig. 9. For a particular set of parameters used, the output function of this system was evaluated at the rate of 100 points per minute, and less than one minute was required to obtain complete response data.

The computer processes discussed are simple logical and algebraic operations when considered independently. Their simplicity permits quick and easy programming for general-purpose digital computer application. Properly sequenced on a high-speed computer, these processes combine to form a powerful tool for the evaluation of both sampled and continuous systems.

References

1. DESIGN OF SAMPLED DATA SYSTEMS BY EXTENSION OF CONVENTIONAL TECHNIQUES, W. K. Linvill, R. W. Sittler. *Report R-222*, July 3, 1953, Massachusetts Institute of Technology, Cambridge, Mass.

2. A GENERAL CARD PROGRAM FOR THE EVALUATION OF THE INVERSE LAPLACE TRANSFORM, C. K. Titus. *Journal*, Association for Computing Machinery, Pittsburgh, Pa., vol. 2, no. 1, pp. 18–27.

3. INTRODUCTION TO NUMERICAL ANALYSIS (book), F. B. Hildebrand. McGraw-Hill Book Company, New York, N. Y., 1956, pp. 443–76.

4. LAPLACE TRANSFORM (book), William T. Thomson. Prentice-Hall, Inc., New York, N. Y., 1950.

Subharmonic Response of the Ferroresonant Circuit with Coil Hysteresis

EGON BRENNER
ASSOCIATE MEMBER AIEE

A SERIES circuit consisting of a capacitance, resistance, and a saturable reactor may, because of the nonlinearity, have a markedly different response from that of the corresponding linear circuit when a periodic voltage is impressed.[1-4] This difference manifests itself in two ways: 1. Jump phenomena occur. 2. The response may contain harmonics which are not present in the impressed voltage. The harmonics may be integral multiples or exact submultiples of the frequency of the impressed voltage. This paper deals with subharmonic response, that is the fundamental frequency of the response is an exact submultiple of the impressed frequency. Specifically, the introduction of an idealized nonlinear inductor makes possible the calculation of the effect of hysteresis in the iron on the response; it also makes possible such calculation of the effect of the impressed wave form. Various types and orders of subharmonic response are predicted and their dependence on two normalized parameters is established theoretically and experimentally.

The basic equation for the series RLC (resistance-inductance-capacitance) circuit is

$$d\psi/dt+Ri+\frac{1}{C}\int i\,dt=v(t) \tag{1}$$

where ψ represents the flux linkages. In addition $i=i(\psi)$ where the functional relationship is not linear. For sufficiently small series resistance and for sinusoidal impressed voltage equation 1 may be reduced to the Duffing equation which follows

$$y''+\alpha y'+f(y)=P\sin\omega t \tag{2}$$

where y represents normalized flux linkages, P normalized voltage amplitude and α is a dissipation parameter.[5,6] The function $f(y)$ represents a (normalized) flux-linkage current relationship. A method which has been successfully used to investigate subharmonic response consists of formulating $f(y)$ as a polynomial by applying curve fitting to the normal induction curve of the iron. Equation 2 is then treated by various approximation methods of the theory of ordinary nonlinear differential equations, e.g., perturbation, Fourier coefficient comparison, and the method of varying amplitude and phase.[7-9]

This approach, however, is not fruitful in the problem at hand because: 1. Hysteresis in the iron cannot be included. 2. Nonsinusoidal periodic wave forms are to be considered as well as sinusoids. 3. Small nonlinearity as required by the formulation of equation 2 is often unrealistic.

A physically meaningful approach was taken by Travis and Weygant, who idealized the normal induction curve somewhat and carried through a piecewise linear analysis with sinusoids applied.[10] Their approximation for the induction curve, shown in Fig. 1, assigns infinite slope to the flux-linkage current curve in the unsaturated region. As a result the capacitor charge can change only during an excursion into saturation; at other times it is "trapped" on the capacitor. This concept of trapped charge is a plausible physical explanation of subharmonic response.

It has been used and extended in the following.

Paper 56-686, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted January 4, 1956; made available for printing May 24, 1956.

EGON BRENNER is with The City College of New York, New York, N. Y.

This paper is abstracted from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Electrical Engineering at the Polytechnic Institute of Brooklyn in June 1955. The author is grateful to Dr. Ernst Weber of the Polytechnic Institute of Brooklyn for his expert guidance and many fruitful suggestions. Special thanks are also due Dr. E. J. Smith for his generous help and encouragement. The assistance of Mr. Ken Lian in furnishing the coils used in the experiments is gratefully acknowledged. The research was supported in part by the U. S. Navy under contract no. *Nonr 835(15)*.

Ideal Nonlinear Inductor

The idealizations which are introduced in the analytical portions of this paper consist essentially of assigning infinite slope to the flux-linkage current curve in the unsaturated region, and zero slope in the saturated region. In addition, hysteresis is included but the series (ohmic) resistance of the coil is neglected.

IDEAL HYSTERESISLESS INDUCTOR

The nonlinearity of an inductor without hysteresis is idealized completely by defining: $i=0$ for $-\psi_s<\psi<\psi_s$ and $d\psi/di=0$ for $|\psi|=\psi_s$. This is illustrated in Fig. 2(A).

IDEAL INDUCTOR WITH HYSTERESIS

For the hysteresis loop corresponding to saturation flux density, Fig. 2(B), define: $i=\pm i_c$ for $|\psi|<\psi_s$ and $d\psi/di=0$ for $|\psi|=\psi_s$. For smaller loops assume that $i_c/i_1=\psi_s/\psi_1$. For minor loops, as illustrated in Fig. 2(C), assume that if ψ has been decreasing from ψ_1 to ψ_2 and if at $\psi=\psi_2$, $d\psi/dt$ becomes positive, then ψ will increase with $i=i_2$ where $\psi_1/i_1=\psi_2/i_2$. This heuristic assumption concerning minor loops will be referred to as the "proportional rule." It is also noted that the hysteresisless inductor as defined in the foregoing is a special case of the inductor with hysteresis for $i_c=0$.

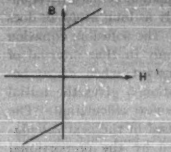


Fig. 1. Induction curve of Travis and Weygant[10]

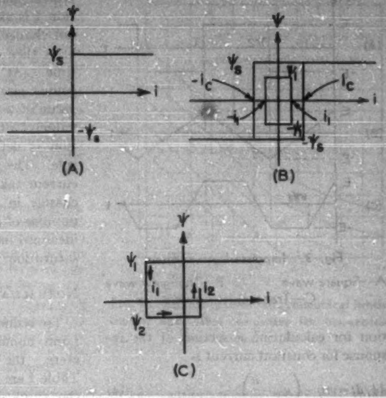


Fig. 2 (right). Idealized coil characteristics

A—Without hysteresis
B—With hysteresis
C—Idealized minor loop

Conditions for Various Responses

The equation governing the idealized circuit is

$$d\psi/dt=v(t)-v_c(t) \tag{3}$$

where $v_c(t)$ represents the capacitor voltage and $v(t)$ is the periodic impressed voltage with half-wave symmetry. Consider, now, an instant of time $t=t_1$ when $d\psi/dt$ is positive so that $v(t_1)>v_c(t_1)$ and the coil is about to enter saturation so that $\psi(t_1)=+\psi_s$. When the coil enters saturation $d\psi/dt$ becomes zero because no further increase in flux linkages is possible. Consequently an impulse of current, $i(t)=C[v(t_1)-v_c(t_1)]\delta(t-t_1)$, will develop to charge the capacitor to the value of impressed voltage. The capacitor voltage will then remain equal to the impressed voltage as long as the impressed voltage increases or remains constant. If the impressed voltage starts to decrease, current in the reverse direction (to discharge the capacitor) will have to flow. Hence, $d\psi/dt$ becomes negative and the counter electromotive force again has a value. One therefore deduces that at the instant of time just before the coil leaves the saturated condition, the capacitor voltage is exactly equal to the impressed voltage.

Since the steady state is to be obtained, these considerations reveal an approach which makes it possible to calculate a typical response cycle in the steady state by starting always with the flux-linkage capacitor-voltage condition as has been stated. It is reasonable to expect the coil to enter the saturated region during the transient for several reasons: 1. The idealization of square loops is expected to be best for the sensitive magnetic materials, which have a high ratio of residual to saturation flux density. 2. Subharmonic response can occur without hysteresis so that the nonlinear effect of saturation must be used for the production of subharmonics. 3. Experimental observation shows that the assumed situation is reasonable.

The method of calculation of response, then consists of starting at the beginning of a half-cycle when $v(t)$ starts to decrease, and solving a sequence of linear equations until the coil again enters saturation. The largest interval for a piece of the solution which can be used in this method is a half-cycle of impressed voltage. The responses which are obtained fall into the following categories under the condition cited, assuming that the calculation is started at $\psi=+\psi_s$ (positive saturation):

1. If $\psi=-\psi_s$ (negative saturation) is reached during or at the end of the first half-cycle, periodicity is established with the fundamental frequency of the response equal to the impressed frequency. Since the second half-cycle will be the exact negative of the first half-cycle ("mirror" symmetry) this response will be termed "symmetrical fundamental" response.

2. If positive saturation is reached at the end of the second half-cycle or before, and negative saturation has not been reached previously then the response is again of fundamental frequency but it will not have mirror symmetry. This response will be termed as "unsymmetrical fundamental."

3. If neither condition 1 nor condition 2 is fulfilled, then more than one cycle of impressed voltage is necessary to establish periodicity and the response is "subharmonic." Subharmonic response may be symmetrical or unsymmetrical depending on whether periodicity is established by entering negative or positive saturation respectively. The subharmonic frequency may be an even $1/(2n)$ or an odd $1/(2n-1)$ submultiple of the impressed frequency.

Method of Calculation

In this section the method of calculation is outlined. Examples of typical calculations are given in Appendix I.

UNSATURATED REGION

Because a calculation is carried through from the time the coil leaves saturation until it enters saturation again, the method of calculating responses consists of solving the appropriate circuit equations with the coil unsaturated. Moreover, except for the transitions from an ascending to a descending portion of a hysteresis loop when $d\psi/dt$ changes sign, the current is constant. A typical equa-

Table I. Normalization and Notation

Physical Quantity	Normalized Quantity
Flux linkages ψ	$y=2\psi/ET$
Capacitor voltage v_c	$u=v_c/E$
Current i	$a=iT/(2EC)$
Time t	$x=2t/T$
Saturation flux linkages ψ_s	$y_s=2\psi_s/ET$ or $Y=f/E$ (f in Kc, E, volts)
Current at coercive force i_c	$a_c=i_cT/(2EC)$
Loop width parameter i_c/ψ_s	$K=a_c/2y_s=i_c(T/2)^2/2\psi_sC$ or $W=1/f^2C$ (f in Kc, C in μfd)

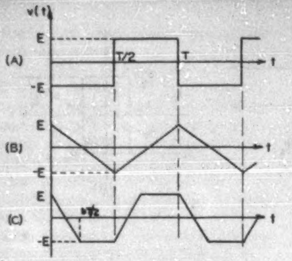


Fig. 3. Impressed wave forms

A—Square wave B—Triangular wave
C—Trapezoidal wave

tion for calculating a section of the response for constant current is

$$d\psi/dt=v(t)-\left(v_{co}-\frac{it}{C}\right) \tag{4}$$

which integrates to

$$\psi-\psi_0=\int_{t_0}^{t}v(\tau)d\tau-v_{co}{}^{(t-t_0)}+\frac{1}{2C}i(t-t_0)^2 \tag{5}$$

where ψ_0 and v_{co} are the initial values for the section of the response which is being calculated. They must be identical to the final values of the preceding calculation.

Because of the minor hysteresis loops it is clear that an equation typified by equation 4 is valid only as long as $d\psi/dt$ does not change sign. The solution, equation 5, is then used only in that interval of time. At the instant when $d\psi/dt$ becomes zero, equation 5 gives the initial value of ψ for the next calculation. The value of i to be used in the next calculation is determined by the proportional rule. The transition to the new value of current takes place in zero time with no change in capacitor charge because an impulse of current cannot develop in the idealized inductor when the coil is not in saturation.

NORMALIZATION

To reduce the problem to the minimum number of variables and parameters, the normalizations shown in Table I are introduced. In this table E is the amplitude of the impressed voltage which has period T. These normalized quantities are not only convenient but have physical significance:

1. $ET/2$, which is used to normalize ψ, is equal to the flux linkages which would be produced if a constant voltage E were applied to the coil directly for a half-cycle.
2. The parameter a is the ratio of the charge on the capacitor due to a current i for a half-cycle to the charge on the capacitor if a constant amplitude E is applied.
3. The parameter K measures the ratio of width to height of the largest hysteresis loop. Note that it is independent of the impressed amplitude E.

Responses of Ideal Circuit

SQUARE-WAVE RESPONSE

The use of square waves for obtaining typical response in linear circuits is extensive because of the validity of superposition. While in nonlinear systems the response to no particular wave form can be considered typical in general, analysis and experimentation using symmetrical square-wave excitation is fruitful in the problem under consideration. The problem of phasing (instantaneous value) in particular is minimized because of the flatness of the square wave. Moreover, the square wave has not only the virtue of leading to simple calculations which yield significant results; it can also be easily produced.

If the symmetrical square waves shown in Fig. 3(A) are applied, the calculation may be started at $t=0$ with $v(t)=-E$, $y=y_s(\psi=\psi_s)$, $u=+1(v_c=+E)$, $a=a_c(i=i_c)$ and carried through sectionally as already outlined. Some details of these calculations are given in Appendix I. The significant conclusion is that a series of inequalities involving only y_s and a_c

Vol.
Pt. 1
1956

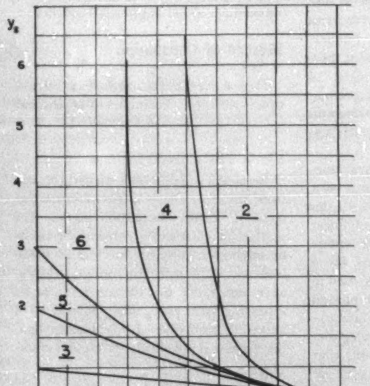


Fig. 4. The $y_s - a_c$ plane for square-wave excitation

Region 1—Symmetrical fundamental response
Region 2—Unsymmetrical fundamental response
Region 3—Symmetrical one-third harmonic
Region 4—Unsymmetrical one-half harmonic
Region 5—Symmetrical one-fifth harmonic
Region 6—Higher orders of subharmonics

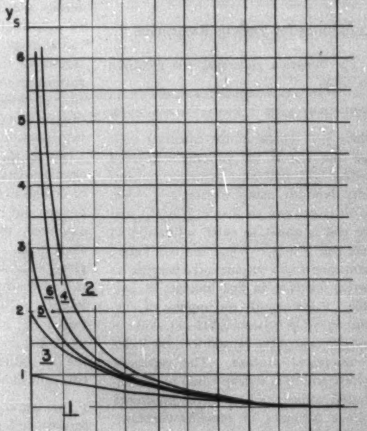


Fig. 5. The y_s—K plane for square-wave excitation. Numbers within regions denote same type of response as in Fig. 4

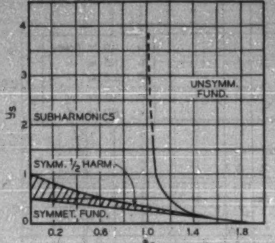


Fig. 6. The y_s-a_c plane for triangular-impressed wave form

or K determine the type of response obtained. These inequalities divide the y_s-a_c plane or the y_s-K plane into regions, each region corresponding to a particular type of response. Fig. 4 shows the y_s-a_c plane and Fig. 5 shows the y_s-K plane. In these illustrations the response regions are identified as far as they were calculated. (In the exceedingly narrow regions of subharmonic response in the y_s-K plane, above $K \approx 1/2$, the regions are shown somewhat simplified. Where the boundaries of these regions are so close together, the indicated order of subharmonic may not occur.)

Conclusions from Figs. 4 and 5

The following conclusions may be drawn from the analysis of the idealized circuits with square-wave excitation, regarding the influence of hysteresis on subharmonic response:

1. Comparison of the analyses with and without hysteresis ($a_c=0$) reveals that with a finite-width hysteresis loop two types of harmonic and subharmonic response are possible, i.e., symmetrical and unsymmetrical, while only the symmetrical responses are possible without hysteresis.

2. The circuit response to square waves is determined by two parameters: y_s, the saturation flux linkages normalized with respect to the half-cycle (applied) voltage-integral; and K, the normalized hysteresis loop width.

3. If the normalized hysteresis loop width exceeds a critical value ($K>2$) no subharmonics can occur. This shows that the energy dissipating feature of hysteresis has the same effect on suppressing subharmonic responses as does ohmic damping.

4. Unsymmetrical responses are due to hysteresis. The even-order subharmonics can also occur only with nonzero width hysteresis loop (for this type of excitation). Moreover, the response regions shown in the y_s-K plane show that the regions of unsymmetrical subharmonic response adjoin the region of unsymmetrical fundamental response. Thus when, for example, the loop width is below the critical value referred to in item 3, if y_s is increased from a very low value (or the amplitude of the applied voltage is decreased from a correspondingly large value), one will observe the following responses in sequence: fundamental, one-third harmonic, one-fifth harmonic, and all other higher odd subharmonics; then the even-order subharmonics in the inverse order, i.e., the one-quarter, the one-half, and finally unsymmetrical fundamental response. The odd-order subharmonics are symmetrical, the even-order unsymmetrical.

5. Within a region on the response plane (y_s-K plane), the subharmonics are stable (because of the idealization of the coil), so that a change of parameter within a region will change the wave form of the response but not its fundamental frequency.

6. As is implicit throughout the analysis (in view of the proportional rule), the minor hysteresis loops are the property of hysteresis, which is responsible for the possibility of even-order unsymmetrical subharmonic response.

Triangular Wave Form

The results obtained with the triangular wave form of Fig. 3(B) impressed are summarized in the response diagram (y_s-a_c graph) of Fig. 6. The calculations in this case were carried only to the point where the similarities and differences between Fig. 6 and Fig. 4 became evident. Note that for this wave form even-order subharmonics may occur with zero-width hysteresis loop. Unsymmetrical response can, however, occur only for finite-width hysteresis loop ($a_c>0$).

Trapezoidal Wave Form

The boundary curves between symmetrical fundamental and other responses for the trapezoidal wave form of Fig. 3(C) impressed, are shown in Fig. 7 for several trapezoids. It is noted that the triangular and square-wave cases are the limiting cases of trapezoids for zero and maximum "flat top" respectively. For this reason it is concluded that the dependence of subharmonic regions in the response diagram on two parameters (y_s and K) is, within the limitations of the idealizations, generally true for periodic wave forms. Thus, for example, the response to sinusoids may be approximated by obtaining the response to a trapezoid. This procedure avoids the difficulty of solving transcendental equations which sinusoids introduce into a piecewise linear analysis.

Theoretical Discussion of Idealizations

The analysis based on the idealizations reveals regions of subharmonic response which depend on the normalized parameters. Although this result can qualitatively be verified in the laboratory, it is appropriate here to discuss each feature of the idealizations from a theoretical viewpoint so that the extent of the validity of the analysis may be judged.

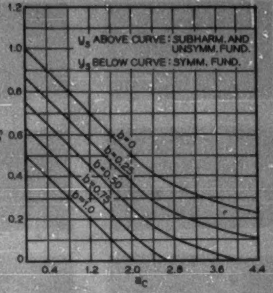


Fig. 7. Boundary between symmetrical fundamental and other responses for trapezoidal wave forms impressed

Series Resistance

Omitting series resistance from the analysis does not influence the validity of the results provided that the resistance is sufficiently small, as smallness is defined in this section.

In the saturated region, the inclusion of series resistance in the analysis shows that the impulse of current which was previously obtained when the coil enters saturation is actually an exponential decay. If the period T of the impressed wave form is large compared to the time constant RC, then the capacitor charge just before the coil emerges from saturation will still equal the (instantaneous) value of the impressed voltage. The same inequality ($T>>RC$) will ensure that the volt drop across the series resistance in the unsaturated condition does not materially affect either $d\psi/dt$ or the capacitor voltage.

Slope in Unsaturated Region

If the assumption of infinite slope of the $\psi-i$ curve in the unsaturated region is relaxed to permit the line to have large but finite slope of value L_0, then the circuit in this section is a linear RLC circuit. Linear analysis shows that if the circuit is taken as very underdamped ($4L_0>>R^2C$) then the change in capacitor charge per half-cycle differs only slightly from the value obtained with the infinite slope assumption, provided also that the impressed frequency is much larger than the natural frequency of the RLC circuit.

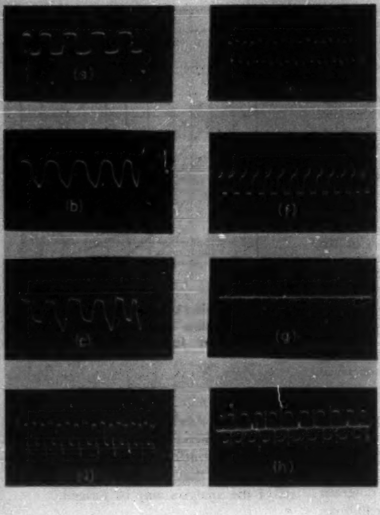


Fig. 8. Current oscillograms obtained by increasing square-wave amplitude impressed on a circuit with 4–79 Mo Permalloy core at 1.5 kc. Capacitance 0.08 microfarad

a—Impressed wave form
b—Unsymmetrical fundamental response
c—One-half harmonic
d—One-quarter harmonic
e—Unsymmetrical one-third harmonic
f—Unsymmetrical fundamental
g—Symmetrical fundamental
h—One-third harmonic (with impressed wave superposed)

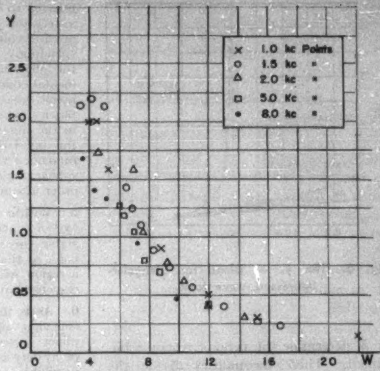


Fig. 9. Square-wave test results: boundary for transition to one-half harmonic

Abrupt Breaks and Slope in Saturated Region

The lack of abrupt breaks in a realistic coil characteristics may be expected to invalidate the quantitative results which are calculated (on the idealized basis) for the higher-order subharmonics, because for these responses the coil would be near the region of the rapidly changing slope for a considerable fraction of an impressed voltage cycle.

The assumption of zero slope in the saturated region is mostly responsible for the approximate nature of the numerical results. Permitting the coil characteristics to have finite but small slope (of value L) again results in a linear *RLC* circuit problem. Now, however, both R and L are small so that this circuit may be either underdamped or overdamped. Consequently the capacitor charge, at the instant just before the coil emerges from saturation, may be larger or smaller than the instantaneous value of the impressed voltage. Hence the corresponding initial condition used in the calculations may be unrealistic. These considerations do not, however, invalidate the type of response diagram expected; rather they modify the relative sizes of the various regions.

Capacitor Charge; Conclusions

With the idealized characteristics, transitions from descending to ascending (or vice versa) portions of the hysteresis loop take place in zero time with no change in capacitor charge. In an actual coil the minor loops will extend into saturation and the transition does not take place with constant flux density. In view of these and the other considerations, the responses of an actual circuit may depart from ideal in these respects:

1. There would be quantitative differences with respect to the width of the response regions.
2. Additional regions (i.e., types or orders of subharmonics) may occur in the realistic case; in particular, unsymmetrical odd orders may be observed.
3. No changes in the general type of response diagram are expected

Experimental Results

The experimental work was carried out to determine the extent to which the type of response depends on the normalized parameters. Since no exact numerical agreement is expected, the parameters Y and W as defined in Table I are used to replace y_s and K respectively.

Square-Wave Tests

Square waves of reasonable rise time and flatness, Fig. 8(A), were applied to an RLC series circuit (in which the coil was wound on a *4-79* Mo Permalloy core) and the voltage was increased from zero ($Y=$ infinite). Initially unsymmetrical fundamental response, Fig. 8(B), was obtained; then the one-half harmonic, Fig. 8(C), occurred. The lowest amplitude of impressed voltage for which the one-half harmonic was stable was recorded for five frequencies and various capacitance values. The results are shown in Fig. 9. The curve formed by the points of Fig. 9 corresponds to the boundary between regions 2 and 4 in Fig. 5 (or Fig. 4).

In Fig. 10 the points of Fig. 9 for the three lowest frequencies are shown as line *1*. Line *2* is the (averaged) curve (for these three frequencies) at which the transition to the one-quarter harmonic occurred. This type of response is shown in Fig. 8(D). Line *3*, Fig. 10, is within the region of an unsymmetrical one-third harmonic, shown in Fig. 8(E), which is not predicted by the ideal analysis. This response, as well as other responses between lines 3 and 4 are discussed in Appendix II. Near line *4* the unsymmetrical fundamental, Fig. 8(F) occurs briefly, then symmetrical responses, Fig. 8(G) and (H), fundamental harmonic and one-third harmonic respectively occur. Further discussion is given in Appendix II.

Sinusoidal Tests

Sinusoidal voltage impressed on the same circuit yielded similar results. The dependence of the response regions on the normalized parameters is most clearly established by the curve shown in Fig. 11. This curve is the transition from unsymmetrical fundamental response, Fig. 12 (B), to the one-half harmonic, see Fig. 12(C). As the voltage amplitude is further increased, the higher-order subharmonics occur in narrow regions and these regions are therefore not clearly defined. For still smaller values of Y the symmetrical one-third harmonic, Fig. 12(D) and (E), and then symmetrical fundamental ferroresonant response Fig. 12(F) occurs.

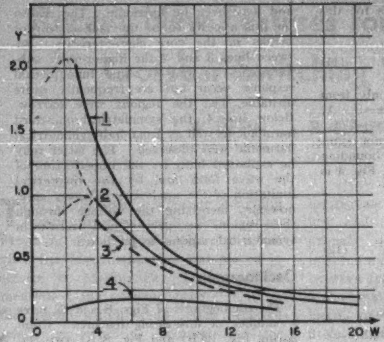


Fig. 10 (left). Square-wave test results: response regions

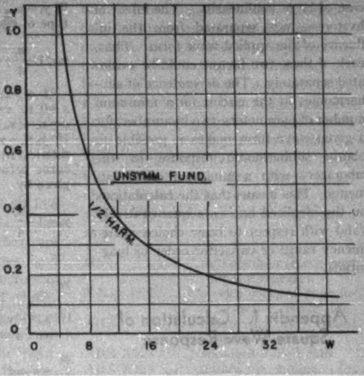


Fig. 11 (right). Sinusoidal test results: boundary for transition to one-half harmonic

CONSISTENCY OF RESULTS

The sequence of occurrence of the various responses were consistent and they were duplicated with a second coil using a Hypernick *V* core. The oscillograms for the latter case were similar, and are omitted in this presentation. It is particularly noted that a response condition, once obtained by varying one parameter in the circuit (e.g., capacitance), could be duplicated by varying another parameter (e.g., frequency) so that the same condition for the normalized parameters was obtained provided, of course, that the response condition was not too near the boundary of a region.

Conclusions

The analysis and the experimental results lead to several conclusions which may be classified in four categories, as follows.

PRODUCTION OF SUBHARMONICS

Subharmonics occur because of the nonlinearity of the coil. When the coil enters saturation, a change in capacitor charge occurs. The release of this capacitor charge in the case of a hysteresisless coil is delayed until the coil again enters saturation. This is the trapped-charge concept of Travis and Weygant.[10] In the case of the coil with hysteresis, the release of the trapped charge is controlled by the unsaturated inductance or the coercive force. A trapped or partially trapped charge may therefore be considered a basic explanation of subharmonics.

The type and order of subharmonic which may occur is determined by three factors: the wave form of the applied voltage, the amplitude and frequency of the applied voltage and hence by the normalized saturation flux linkages, and the normalized hysteresis loop width.

The type and order of subharmonics for a given circuit may be summarized (for a given applied voltage wave form) in a response diagram which is a plot of regions in the y_s (or Y) (normalized saturation flux linkage)$-K$ (or W) (normalized loop width) plane. If the idealizations of this analysis are used, the response diagram shows the conditions for which a given type and order of subharmonic may occur. With actual coils the size and number of the response regions are modified. The series resistance of the circuit must be small or subharmonics cannot occur.

INFLUENCE OF COIL HYSTERESIS

Without hysteresis only symmetrical responses occur. Hysteresis accounts for unsymmetrical responses of both fundamental and subharmonic variety. If the hysteresis loop width exceeds a critical value, no subharmonics can occur.

INFLUENCE OF IMPRESSED WAVE FORMS

Symmetrical even- and odd-order subharmonics are predicted with periodic wave forms having quarter-wave symmetry, half-wave monotonicity and zero average value except for the symmetrical square wave. For the latter case, only symmetrical odd orders are predicted.

A slight change in wave form (such as changing from one trapezoidal type to another of the same frequency but slightly longer flat top) produces only correspondingly small changes in the response diagram. The actual current wave forms within a response type, however, are sensitive functions of the voltage wave forms.

VALUE OF ANALYSIS WITH IDEALIZED CHARACTERISTICS

The work presented in the foregoing shows that the sectional linearization, so as to emphasize the nonlinearity and consequent idealization of coil characteristics, is of value in furthering a basic understanding of subharmonic response from the circuit point of view in the realistic

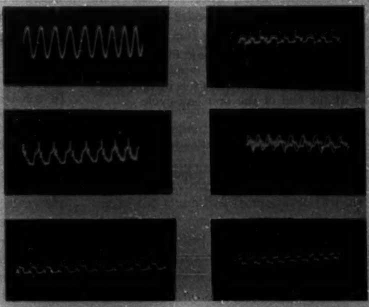


Fig. 12. Current oscillograms obtained by increasing sine wave amplitude in the test circuit at 900 cycles per second. Capacitance 0.04 microfarad

a—Impressed wave form
b—Unsymmetrical fundamental response
c—One-half harmonic
d—One-third harmonic
e—Same as (d) with impressed wave form superposed
f—Symmetrical fundamental response

case of large nonlinearity. The effect of hysteresis was separated from the influence of the applied wave form. Thus, each of these two factors could be evaluated separately. The dependence of subharmonics in the circuit on a minimum number of parameters, two in number, for a given wave form makes it possible to obtain subharmonic response in the laboratory with a minimum of adjustments. This means that the calculations on the idealized basis are quantitatively valid with respect to basic choice of frequency range or capacitor order of magnitude.

Appendix I. Calculation of Square-Wave Response

Calculation Without Hysteresis

The wave form of Fig. 3(A) is applied to the circuit, series resistance ignored, and the coil characteristics of Fig. 2(A) used. Starting with $y=+y_s$, $u=+1$ and $v(0)=-E$, the normalized circuit equation for the first half-cycle of impressed voltage is

$$dy/dx=-2;\ 0<x<1 \tag{6}$$

Integrating

$$y=y_s-2x \quad 0<x<1 \tag{7}$$

so that symmetrical fundamental response occurs for $y_s \leq 1$. If $y_s>1$ then, during the second half-cycle $dy/dx=0$, so that subharmonic response occurs. During the third half-cycle, let $x'=x-2$

$$y=y_s-2-2x',\ 0<x'<1 \tag{8}$$

Hence, if $1<y_s\leq 2$ "negative" saturation $(y=-y_s)$ is reached during or at the end of the third half-cycle, so that symmetrical one-third harmonic response results. In general if $m-1<y_s<m$, the $1/(2m-1)$ harmonic occurs (with m as an integer).

Calculation With Hysteresis: First Half-Cycle

The initial conditions are the same as without hysteresis. Hence for the first half-cycle

$$\frac{dy}{dx}=1-u;\ u=1-a_cx \text{ or } \frac{dy}{dx}=-2+a_cx \tag{9}$$

If $a_c<2, dy/dx$ is monotonic for the first half-cycle and equation 9 applies up to $x=1$. In this case equation 9 is integrated, and at the end of the first half-cycle

$$y(1)=y_1=y_s-2+\frac{a_c}{2} \tag{10}$$

Symmetrical fundamental response occurs if y_1-y_s, so that the conditions for this type of response are

$$y_s<1-\frac{a_c}{4};\ a_c<2 \tag{11}$$

If $a_c>2$, equation 9 applies only from $x=0$ to $x=x_a=2/a_c$, when $dy/dx=0$. At $x=x_a$ $u(x_a)=-1$ and $y(x_a)=y_a=y_s-2/a_c$. If $y_a\leq -y_s$ the response is again symmetrical-fundamental so that the bounding curve between region 1 and 3 in Fig. 4 is given by

$$\begin{aligned} y_s&=1-\frac{a_c}{4} && a_c<2 \\ y_s&=\frac{1}{a_c} && a_c>2 \end{aligned} \tag{12}$$

If $y_a>-y_s$, define $n_a=y_a/y_s$ and set $x'=x-x_a$. For the remainder of this half-cycle

$$\frac{dy}{dx'}=n_aa_cx' \tag{13}$$

where $-n_aa_c$ is the value of a adjusted in accordance with the proportional rule. If $n_a>0$, y increases. If $n_a<0$ equation 13 indicates decreasing y with positive current, an impossible condition (with the idealized characteristics) so that $y=y_a$ (with $a=0$) until the impressed voltage reverses. In either case the flux linkages will be increasing during the second half-cycle at least as rapidly as during the first half-cycle so that the condition $y_s>1/a_c$ for $a_c>2$ results in unsymmetrical fundamental response.

Subsequent Half-Cycles

One need only consider the case

$$y_s>1-\frac{a_c}{4};\ a_c<2 \tag{14}$$

since the other cases are completed in the preceding paragraph. Now define $n_1=y_1/y_s$ where y_1 is given in equation 10. During the second half-cycle $(x^1=x-1)$

$$\frac{dy}{dx'}=1-(1-a_c[1+n_1x'])\quad 0<x'<1 \tag{15}$$

The procedure now follows the treatment of equation 9, and the next curve desired is the boundary between region 2 and 4 in Fig. 4. In a similar manner the other curves in Figs. 4 and 5 are calculated.

Appendix II. Discussion of Experimental Results

Discussion of Fig. 10

In constructing line 1 of Fig. 10, only the results of the frequencies from $1-2Kc$ were used because of the variation of coercive force with frequency. Near the boundary of each region, instability occurs. The response there is not clearly defined and may even be aperiodic. The reason for this appears to be the lack of abrupt breaks in the core characteristics. Between lines 3 and 4 the higher-order subharmonics as well as some fundamental response occur but are frequently quite unstable, i.e., the regions were narrow. Below line 4, the symmetrical one-third harmonic as well as a one-half symmetrical harmonic were observed. The latter may be explained both by the imperfection of the wave form and by the theoretical limitation of the analysis. In every case, however, increasing the voltage through these regions (i.e., decreasing Y) results in symmetrical response as predicted.

Oscillograms

The oscillograms of Figs. 8 and 12 have a number of interesting features. Comparing Fig. 12(B) and Fig. 8(B), two unsymmetrical fundamental responses, it is noted that while the type of response is the same in the two cases, the actual wave forms differ. The same remark applies to Figs. 12(C) and 8(C). The type of observation motivates the statement concerning the wave forms in the second paragraph of the conclusion "Influence of Impressed Wave Forms."

Examination of the oscillograms shows also that in each case of subharmonic response, excursions into saturation take place and that in the intervening regions the current is almost constant except for much smaller excursions into saturation. These observations tend to enhance the value of the idealizations.

References

1. CONCERNING A NEW FREQUENCY METER OF THE SIEMENS AND HALSKE CORPORATION, **O. Martienssen.** *Electrotechnische Zeitschrift*, Wuppertal-Elberfeld, Germany, Feb. 24, 1910, pp. 204–06.

2. CONCERNING A NEW RESONANCE PHENOMENON IN A-C CIRCUITS, **O. Martienssen.** *Physikalische Zeitschrift*, Leipzig, Germany, vol. 11, 1910, pp. 448–60.

3. A STATIC FREQUENCY DIVIDER, **J. Fallou.** *Revue Generale d'Electricite*, Paris, France, vol. 19, 1926, p. 987.

4. SUBHARMONICS IN FORCED OSCILLATIONS IN DISSIPATIVE SYSTEMS. **P. O. Pederson.** *Journal Acoustical Society of America*, New York, N. Y., vol. 6, 1935, p. 227.

5. HARMONIC AND SUBHARMONIC RESPONSE OF THE DUFFING EQUATION. **M. E. Levenson.** *Journal of Applied Physics*, New York, N. Y., vol. 20, Nov. 1949, p. 1045.

6. NONLINEAR VIBRATIONS (book), **J. Stoker.** Interscience Publishers, New York, N. Y., 1950.

7. RESONANCE, **J. Ludeke.** *Journal of Applied Physics*, New York, N. Y., vol. 13, 1942, p. 418.

8. PREDOMINANTLY SUBHARMONIC OSCILLATIONS, **J. Ludeke.** *Ibid.*, vol. 22, Nov. 1951, p. 1321.

9. SUBHARMONIC OSCILLATIONS IN NONLINEAR SYSTEMS, **C. Hayashi.** *Ibid.*, vol. 16, 1945, p. 105.

10. SUBHARMONICS IN CIRCUITS CONTAINING IRON-CORED REACTORS, **Irven Travis, C. N. Weygandt.** *AIEE Transactions* (*Electrical Engineering*), vol. 57, Aug. 1938, pp. 423–31

Recent New Features for the No. 5 Crossbar Switching System

J. W. DEHN
NONMEMBER AIEE

R. E. HERSEY
NONMEMBER AIEE

THE NUMBER 5 crossbar switching system was described by Korn and by Ferguson[1] in 1950, just a year and a half after the first installation was placed in service at Media, Pa. These authors mentioned the growing need "...for a dial system which would better meet the telephone switching requirements for areas on the outskirts of metropolitan cities and at the same time care for medium-to-large sized offices in other areas." Since that time, the basic switching plan of this system has proved to be so flexible and adaptable that it is finding an ever-expanding application in many new combinations of switching requirements. After a brief review of the basic switching plan described in more detail in the previous article, this paper will cover some of the new features since added to the system.

No. 5 Crossbar System: Basic Switching Plan

Like its forerunner, the No. 1 crossbar system,[2] the No. 5 is an all-relay, and relaylike, switch system. Fig. 1 indicates the major circuits of the system and the manner in which they are interconnected. The blocks called line link frames and trunk link frames represent the crossbar switch frames each having columns of ten primary and ten secondary switches. The quantities of such frames depend upon the number of lines to be served, their calling rates, and average conversational holding time. A single switching unit, usually called a marker group, having four line link frames and two trunk link frames, is the minimum size usually provided and may serve about 2,000 lines. Other marker group sizes range up to 40 line link frames and 20 trunk link frames and may serve up to 40,000 numbers. Arrangements are being made to increase the frame limits to 60 line link and 30 trunk link frames. The primary and secondary switches are interconnected between switches and between frames by links and junctors in a systematic "fanning out" and "interlaced" pattern so that there are always at least ten paths between any line on the left and each trunk on the right. Customers' lines are wired to the verticals of the columns of switches on the left and the various circuits to which customers can be connected are wired to the horizontals of the column of switches on the right. Included on the right, in addition to all the various types of trunks, are the originating registers in which are recorded all the digits of each call as the customer dials them. Tandem-type and toll-type incoming trunks also have an appearance on the verticals of the line link primary switches on the left, in order that calls may be switched through to outgoing trunks. Incoming calls are served by incoming registers capable of receiving the type of pulsing impressed at the far end. On the other hand, certain outgoing calls require outgoing senders capable of delivering the type of pulsing which the far end is arranged to receive. Incoming registers and outgoing senders are connected to their respective trunks through crossbar switch links. All of the major circuits have connections to markers through multicontact relay circuits called connectors. A marker is used for establishing the connections for each call. Details are covered in the Korn-Ferguson paper.[1]

New Features

Some features added to provide improved handling of local and short-haul traffic will be described here.

PRETRANSLATORS

It will be recalled that after a customer has dialed his call into a register, a marker is utilized to complete the desired connection. As customers are permitted direct dialing for more and more of their calls, more variation in the number of digits dialed may be encountered. It therefore becomes desirable to automatically determine completion of dialing without introducing a delay of several seconds awaiting additional digits which may not be dialed. Accordingly, a circuit called a pretranslator[3] has been made available to the originating registers on a selectable basis. This circuit provides means for determining the total number of digits by inspecting the first three (code) digits dialed. The variety of digits which may be dialed can be represented as follows:

O................	zero operator (recognized in the register).
XII or IIX........	service codes such as information, repair, etc.
ABX.............	unintentionally dialed unused local codes.
ABX XXXX......	standard 7-digit calls.
ABX XXXX P...	a few remaining calls with party letters.
XOX.............	unintentionally dialed unused area codes.
XOX ABX XXXX.	direct distance dialed calls.
XOX ABX XXXX P.	direct distance dialed calls with party letter.

As soon as three digits are dialed, they are passed to the pretranslator which can then determine whether 3, 7, 8, 10, or 11 digits are required. Thus, the register can proceed to call the marker immediately when the proper number of digits has been received. Except when O is dialed, the pretranslator functions on all calls. Its holding time is about 150 milliseconds. Usually two, mounted on a single bay 23 inches wide, will be enough for one central office unit; see Fig. 2.

SEPARATE DIAL TONE AND COMPLETING MARKERS

As described in the previous paper, the control of the switching elements is centered in a few common control circuits called markers, which are rapid in operation. A new feature introduced since the previous paper isolates the functions of the marker which control selection of and connection to the originating registers, and places them in a circuit called a "dial tone" marker which has no other function. It will be recalled that in establishing a connection from a line to an originating register, the marker has but one fixed set of functions to perform: It identifies the calling line, chooses an idle originating register associated with one of the trunk link frames, selects a set of links and junctors between the line and register, and, by operating the associated crossbar switch elements, connects the originating line to the register which together returns dial tone to the customer.

Paper 56-667, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted November 21, 1955; made available for printing May 9, 1956.

J. W. DEHN and R. E. HERSEY are with the Bell Telephone Laboratories, Inc., New York, N. Y.

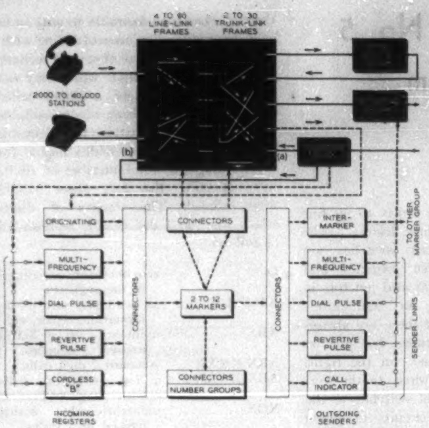


Fig. 1 (left). No. 5 crossbar switching system block diagram. Connection (a) is for terminating connections, and connection (b) is for tandem connections

All other control features remain in another circuit called the completing marker. In this way the unused completing features of the marker are not withheld from use while the dial tone function is activated, and vice versa. This increases marker efficiency, thus reducing over-all marker cost.

Enlarging Over-all Capacity

The limiting parameters of No. 5 crossbar offices have been enlarged as follows:

1. Previously limited to two 10,000-directory-number series and two central office names in each series, arrangements now permit 40,000 directory numbers and six central office names with any desired assignment of each 1,000 numbers to any office name not to exceed 10,000.

2. As a result of this added flexibility, variable calling rates, population distribution, geographic boundaries, etc., many more differences of service and rate treatment become imperative. This is the case especially where a single switching unit serves several areas separated by geographic boundaries. In such cases it is possible that each of the six office names may represent a community sufficiently different from the others to warrant a different rate structure. Accordingly, the number of rate and service classes has been increased from 30 to 60.

3. The customers in some of the areas where No. 5 crossbar will be used have such high calling rates and holding times that the limit of 40 line link frames and 20 trunk link frames will permit only about 6,000 lines to be assigned. This would be wasteful of office names and would require the uneconomical introduction of a second marker group for additional lines. On the other hand, the traffic density is sometimes low enough to allow the number of assigned directory numbers to fall in the areas between 10,000 and 20,000 or between 20,000 and 30,000, or between 30,000 and 40,000. In each case enlarging the array of line link frames and trunk link frames up to 60 and 30 respectively will permit filling out each office code to approximately 10,000 numbers. This increase in number of frames is being provided.

Intermarker Group Operation

When the capacity of one marker group is exceeded, it is frequently desirable to serve additional lines by providing a second marker group in the same building. Switching of calls from one marker group to the other could be handled by the same out-trunking and pulsing methods used when the marker groups are in separate buildings. However, a combination of outgoing and incoming trunk circuits can be provided which is much simpler when both parts are adjacent. Such trunk circuits are called intermarker group trunk circuits. Also, it becomes quite simple to pass information laterally from one marker to another without outpulsing the digital information. This is accomplished by providing "intermarker group sender" circuits associated with intermarker group trunk circuits. Such a sender receives the digital information from a marker of the originating group and passes the same digital information directly into a marker of the other group, as from an incoming register; see Fig. 3. The operation is very rapid and therefore a small group of these senders will handle many calls. A similar group handles calls in the reverse direction. Another type of intermarker group trunk circuit, functioning similarly, operates so that calls originating in one marker group may be switched through the other

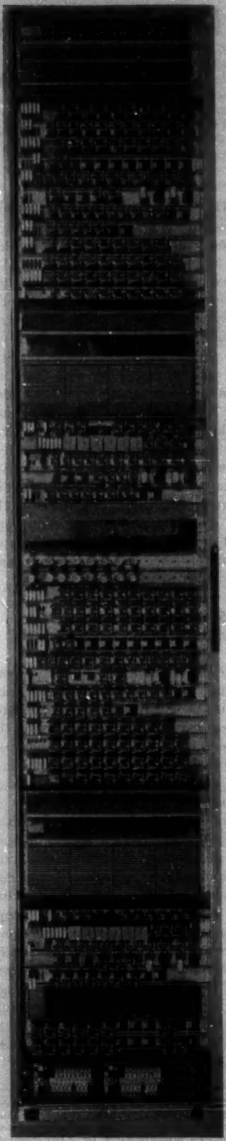


Fig. 2. Two pretranslators on one frame

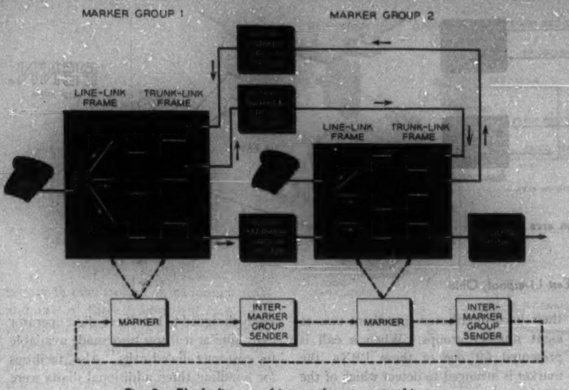


Fig. 3. Block diagram of intermarker group trunking

marker group to any of the outgoing trunks in that marker group. Also provided are the reverse of this, i.e., intermarker group trunk circuits which permit incoming trunks in one marker group to switch calls through to customers in the other marker group.

Junctor Circuit Operation

An interesting arrangement of combined circuitry permits certain types of calls which require special types of supervision and/or control features to be served by a comparatively small group of circuits having the necessary special features. As an example, coin customers require trunk circuits which have various kinds of coin test, coin collect, and coin return features available. For non-coin customers the trunk circuits can be much simpler. Usually the coin traffic is comparatively low and also does not usually coincide with the peaks of non-coin traffic. To route the coin traffic over the non-coin trunks would require that each of a fairly large group of trunks to every destination be provided with both coin and non-coin features. Alternatively two subgroups of trunk circuits, one for coin and the other for non-coin could be provided. To provide the coin features and still realize the economies of a single group of non-coin trunks to each destination, a small special trunk group having the coin features can be provided. This group appears not only as outgoing trunks on the trunk link frames but again as incoming tandem-type trunks on the line link frames; see Fig. 4. Such combined circuits are called junctor circuits. A regular outgoing sender and an incoming register is utilized on each call. The incoming register after receiving the digital information from the sender requests a second marker to route the call through the switches again to one of the non-coin outgoing trunks. Therefore it is neither necessary to provide special features in all outgoing trunk circuits nor to provide duplicate groups of outgoing trunk circuits, having the necessary features, to every destination.

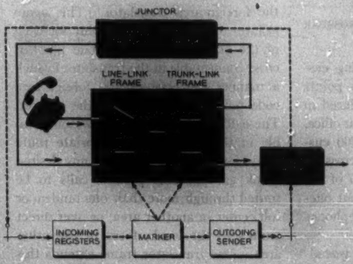


Fig. 4 (left). Block diagram of junctor operation

Marker Pulse Conversion[4]

A feature which utilizes the common control circuits of the system without actually setting up calls through the switches is made use of in connection with operator trunks at a switchboard in the same building; see Fig. 5. In this case the operator selects an outgoing trunk to the desired destination in the switchboard multiple even though the required pulsing to the distant end is not the same as that which the position provides. For example, the position may be provided with multifrequency keying but the distant circuit requires dial pulses. In this case the trunk circuit selected will be split and the cord end will have access to the No. 5 crossbar incoming link and register of the multifrequency type. After the operator keys the digital information into the register, it will call upon a marker to pass the digital information to one of the No. 5 crossbar senders having the appropriate kind of pulsing, and attach the sender to the split trunk for outpulsing. The incoming register and marker are dismissed and, after outpulsing, the sender is released and the talking connection is established through the trunk circuit. No switching path through the line link and trunk link frames is used on this type of connection.

Improved PBX Features

Number group frames are provided, one for each thousand numbers, to indicate to the marker the exact location of each called line and the type of ringing current to apply on party lines. For the usual PBX (private branch exchange) the directory number and all associated lines will appear in one number group.

Where there are large and/or important PBXs it becomes desirable to provide for distribution of the lines over more than one number group. This divides the traffic load to large PBXs so that a given number group will not be called upon for more than its share of the total office load. Some PBXs, such as those handling ticket reservations, provide only as many lines as there are agents. Thus, in some cases when all lines are busy, hundreds of calls may still be offered. For this reason it is

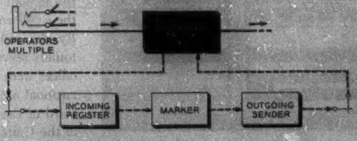


Fig. 5. Block diagram of marker pulse conversion

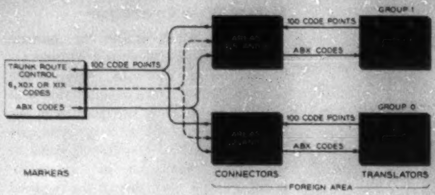


Fig. 6 (above). Block diagram of foreign area translator

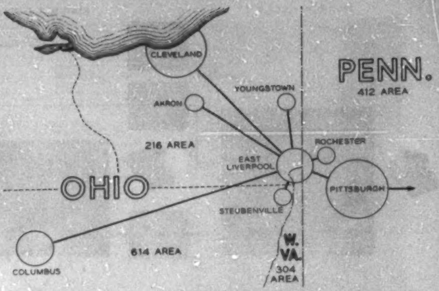


Fig. 8 (right). Foreign area translator used in East Liverpool, Ohio

desirable to be able to recognize such PBX groups as busy without actually entering the number group. New features of this type have been designed into a circuit called a PBX allotter; one is assigned to operate with each marker. With this circuit, ten different PBXs may have their lines distributed into from two to eight number groups. When a call is presented for one of these PBXs, the marker is arranged to detect which of the number groups contain idle lines and then it places the call in one of the number groups which that marker prefers. Different markers have different preferences but each can reach any line. In this way the markers may complete calls to the same PBX without delaying one another.

Fig. 7. Foreign area translator frame

Features Added for Direct Distance Dialing, Tandem and Toll Use

To a degree greater than was earlier anticipated, No. 5 crossbar offices have been located at points where tandem and toll switching is required. Certain uses include directing codes to allow calls to be routed into and out of other switching areas. Accordingly, for tandem and toll use, outgoing senders, incoming registers, and markers were originally arranged to handle a total of ten digits (plus an eleventh where the fast-disappearing party letters were involved). While early use of 10-digit switching was confined to calls handled by operators, the possible extension of 10-digit dialing to customers was anticipated. Some added features which permit customer direct distance-dialing and expand the usefulness of this system in certain tandem and toll applications are described in the following.

TEN-DIGIT DIALING FOR NO. 5 CROSSBAR CUSTOMERS

Late in 1951, a trial permitting customers to dial distance calls to points throughout the nation was introduced in the Englewood, N. J., No. 5 crossbar office. By the end of 1955 about 450,000 customers served by over 50 No. 5 crossbar offices were able to dial calls to 17 of the national areas[5] which include about one-third of the total Bell System telephones in the United States.

Since the No. 5 crossbar system was already designed to handle 10-digit operator calls, similar features were made available for customer-dialed calls. Also, facilities for handling three additional digits were added to the accounting equipment.[6]

FOREIGN AREA TRANSLATOR[7]

On calls to other numbering areas the marker normally recognizes only the area code (XOX or XIX where X represents any value from 2 to 9) to determine the routing. Thus, all calls to a given area may be routed over only a single tandem or toll trunk group. This restriction may result in indirect routing and inefficient use of outside plant. With the introduction of foreign area translation, the area code enables the markers to obtain access to a translator representing a particular called area, thus permitting 6-digit translation; refer to Fig. 6.

The first three digits of all calls which reach a marker are translated in the marker by a relay tree circuit.[8] Through the relay tree a code point is isolated for each three digits and is normally connected to a route relay for directing the course of that call. In the case of XOX or XIX area code digits, the code point may be wired to a connector which selects a foreign area translator serving three areas. The code point also indicates which of three areas is called. The next three digits (office code) are then carried through the connector into a relay tree in the foreign-area translator. This again isolates a particular code point for each of the office codes. This code point is cross-connectable in the translator in such a manner that one of 100 foreign-area code points is indicated to the marker. These 100 code points are now connectable in the marker to appropriate route relays for routing the call over any desired trunk group. This permits calls to be routed through more than one tandem or toll center in another area, or over direct trunks to the local offices in the other area. One translator frame extends this

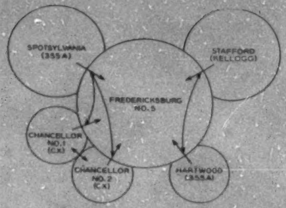

SPOTSYLVANIA
(355A)
STAFFORD
(KELLOGG)
FREDERICKSBURG
NO. 5
CHANCELLOR
NO. 1
(CX)
CHANCELLOR
NO. 2
(CX)
HARTWOOD
(355A)

Fig. 9. Tandem screening

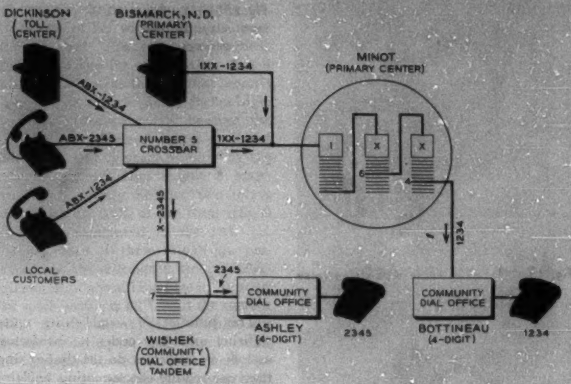

DICKINSON
(TOLL CENTER)
BISMARCK, N. D.
(PRIMARY CENTER)
MINOT
(PRIMARY CENTER)
1XX-1234
ABX-1234
ABX-2345
NUMBER 5
CROSSBAR
1XX-1234
ABX-1234
1
X
X
6
4
X-2345
1234
LOCAL
CUSTOMERS
2345
7
COMMUNITY
DIAL OFFICE
COMMUNITY
DIAL OFFICE
WISHEK
(COMMUNITY
DIAL OFFICE
TANDEM)
ASHLEY
(4-DIGIT)
2345
BOTTINEAU
(4-DIGIT)
1234

Fig. 11. Application of code conversion

service to one, two, or three foreign areas and two of these frames will care for a maximum of six foreign areas. For safety reasons each frame is arranged with duplicate translators. The foreign-area translator frame is shown in Fig. 7.

The East Liverpool office in Ohio has been chosen to illustrate this feature. As indicated in Fig. 8, the East Liverpool office is located at the boundary of three states, Ohio, Pennsylvania, and West Virginia, serving customers in all three states. Although customers served by this office actually reside in three different numbering plan areas, they are all considered to be in the Ohio 216 area, or their "home" area.

It will be noted that there is more than one route into adjacent numbering areas. About 40 community dial offices and tributaries are reached through Columbus in the Ohio 614 area, and some eight or ten tributaries are reached through Steubenville. In the Pennsylvania 412 area there are trunk groups to Rochester and Pittsburgh. Rochester is another No. 5 crossbar office, while trunks to Pittsburgh handle customer-dialed traffic to Pittsburgh offices and through traffic to New York, Philadelphia, Washington, Chicago, Detroit, and other cities. There are, of course, other routes into these areas not shown in Fig. 8. Both local customers' calls and calls originated by customers or operators at distant points which are switched through East Liverpool will utilize the foreign-area translator.

TANDEM SCREENING

When the No. 5 crossbar system serves as a tandem switching point for surrounding community dial and step-by-step offices, means must be provided to limit the direct-dialing access of these offices. Unlike common control systems, such offices are not readily adaptable to making this discrimination at the originating office, and therefore it must be made at the tandem point. This is referred to as tandem screening.

Fredericksburg, Va., is a typical case in point; see Fig. 9. Here there are five small step-by-step offices using the No. 5 crossbar office at Fredericksburg as a tandem point. Customers in all five are permitted to dial Fredericksburg numbers. Chancellor No. 1 and Chancellor No. 2 customers may also dial Spotsylvania numbers, and vice versa. On the other hand, Hartwood and Stafford customers may dial each other but are not permitted to dial beyond Fredericksburg. This is accomplished by providing class signals in the incoming trunk circuits from such offices. These signals are passed through the incoming link and incoming register to the marker where class of service distinction is made so that calls coming in over one group of trunks are extended only to certain offices. Facilities have been made available for providing as many as five screening patterns.

CODE CONVERSION

Code conversion is another feature originally made necessary by the existence of the step-by-step system in the switching network. However, new uses are being found for it. To illustrate the need for arbitrary codes see Fig. 10. Suppose an operator in one toll center

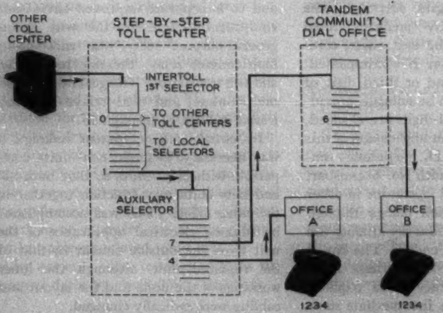

OTHER
TOLL
CENTER
STEP-BY-STEP
TOLL CENTER
TANDEM
COMMUNITY
DIAL OFFICE
INTERTOLL
1ST SELECTOR
0
TO OTHER
TOLL CENTERS
TO LOCAL
SELECTORS
1
6
AUXILIARY
SELECTOR
7
4
OFFICE
A
OFFICE
B
1234
1234

Fig. 10 (left). Illustration of arbitrary codes

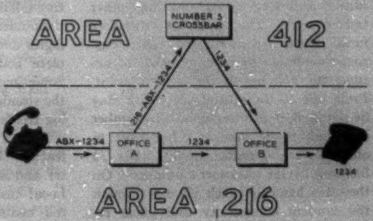

AREA 412
NUMBER 5
CROSSBAR
216-ABX-1234
1234
ABX-1234
OFFICE
A
1234
OFFICE
B
1234
AREA 216

Fig. 12 (above). Diagram showing the use of prefixing digits

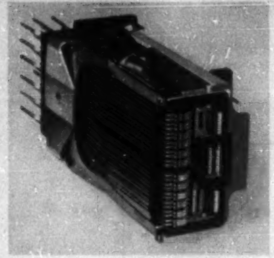


Fig. 13. General-purpose wire spring relay

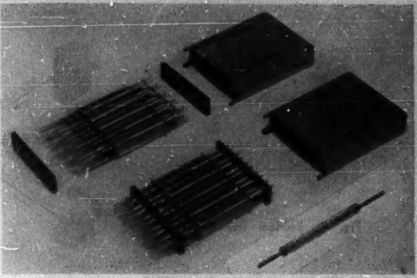


Fig. 15 (right). Five reed relays in one package

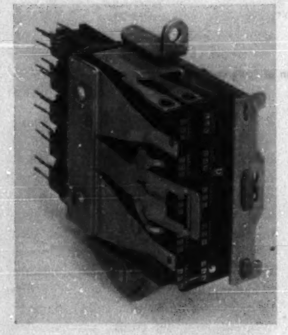


Fig. 14. Wire spring multicontact relay

wishes to place a call to a customer in office *A* of another toll center reached through a step-by-step intertoll train. The incoming trunk from the originating operator terminates in an intertoll first selector. Levels 2 to 9 of this selector are used to reach numbers in the local step-by-step office. Levels 1 and 0 are available for through switching. It is common practice to use level 1 for tributaries and level 0 for reaching other toll centers. The operator therefore must prefix a code to the called customer's number in order properly to actuate the switches at the tandem point. From one to three arbitrary digits may be required, depending upon the number of points to be reached. For calls to office *A*, the operator would prefix *14* followed by the customer's number. On the other hand, to reach office *B* through the tandem community dial office requires an additional digit. Thus, the operator would dial *17* to get through the toll center switches, and *6* to get through the tandem community dial office, followed by the customer's number.

The practice of establishing routes utilizing arbitrary codes is satisfactory as long as operators do the dialing since they can readily check routing bulletins. However, when customers do the dialing, routing bulletins would not be practical. Several rather extensive intertoll dialing networks based on the use of arbitrary codes have already been established. For customer dialing, all offices will be reached on a 2 to 5 numbering basis, each office having a standard ABX office code followed by four numerics. Rearranging switch multiples to use the standard office codes instead of arbitrary codes would be prohibitively expensive. Since this network, based on arbitrary codes, has already been established for operator dialing, it appears reasonable to provide means for code conversion at the last common control entry point into the arbitrary code network.

Code conversion is actually a combination of digit deletion and digit prefixing. The No. 5 crossbar system has therefore been arranged to prefix one, two, or three arbitrary digits ahead of the digits received. In addition, from one to six of the digits already dialed may be deleted. This means that the outgoing dial pulse sender may outpulse the prefixed arbitrary digits and then skip any or all of the following six digits, outpulsing the remainder. Thus, by prefixing three digits and deleting the next three, complete code conversion is accomplished. Sometimes, the second or third digit of the standard code will be suitable, providing the first or second digits are changed.

Fig. 11 illustrates an application of this feature. Bismark, N. D., is a primary center and is equipped with No. 5 crossbar. Local customers, and operators in other toll centers, will reach offices through Bismark No. 5 crossbar by dialing the standard ABX office codes. The No. 5 crossbar office will convert these ABX codes into the arbitrary codes required for switching through intermediate step-by-step points. A call to Bottineau requires an arbitrary *154* to get through the Minot step-by-step primary center. Calls to Ashley require an arbitrary *7* to get through the tandom community dial office at Wishek. There are several others but this suffices to illustrate the principle.

While not so apparent, the need for being able to prefix up to three digits on a multifrequency basis may still be quite common; see Fig. 12. Suppose a customer in office *A* makes a call to office *B* in his home area; he will dial *ABX-1234*. If the direct trunk group to office *B* is busy, an alternate route through a tandem office in the adjacent area may be used. In this case it would be necessary to prefix the home area code ahead of the ABX office code in order to inform the tandem switching equipment that it must route the call back into the adjacent area. In this case the area and office codes are deleted at the tandem point, since there are direct trunks from the tandem to the completing office.

General System Improvements

In addition to the new switching features already described, the No. 5 crossbar system has been redesigned to make use of available new apparatus components, and to incorporate improved circuit and equipment designs, as the wire spring general-purpose relay, the wire spring multicontact relay, the dry reed relay, and the transistor. The general-purpose, multicontact, and reed relays are illustrated in Figs. 13 through 15.

In circuit and equipment redesign it was necessary to make new units compatible with older units so that both old and new units may function together in the same office. This was accomplished by making electrical appearance of the unit from the outside similar to that of the replaced unit, although the inner workings of the units and the layout and cabling were radically changed.

Vol. 7
Pt. 1
1956

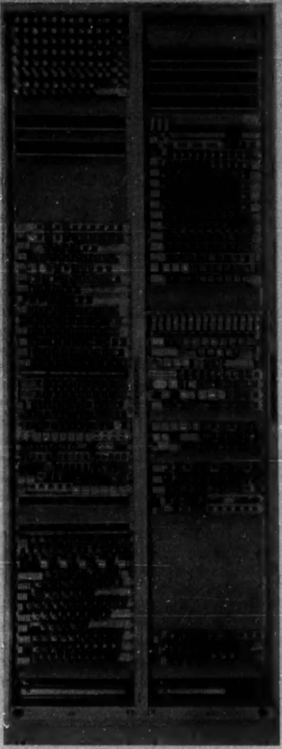


Fig. 16. The old dial tone marker

Typical Redesigns

A few typical redesigns of large, medium, and small units will be described. Figs. 16 and 17 show the old and new dial tone markers. A comparison of these two markers is shown in Table I.

Figs. 18 and 19 show the old and new originating register frames. The circuit has been redesigned to eliminate 14 of the general-purpose relays. In addition, the dry-reed type of relay has been substituted for the more conventional flat-type. A dry-reed relay package containing five coils each actuating two contacts replaces five of the older relays, and occupies a space not much larger than one of its predecessors. This permits eight registers to be mounted on a 2-bay frame which accommodated only five of the older variety. The redesign reduces the current drain by about 35% (per cent).

One of the many small redesigned units is shown with the replaced unit in Fig. 20. This is one of the widely used incoming trunk circuits. The older arrangement mounts two circuits on a 3-plate unit 6 inches high, whereas the new unit mounts one circuit on a single plate 2 inches high. This could be done partly because the wire spring slow-release relays need not have the protecting space required by the older relays to reduce magnetic interference, and partly because two relays (approximately 20% of the total relays) have been eliminated in the redesign. The functions of these relays are performed by the redesigned ringing selection switch at a small increase in its price. The space saving is 33%, and the current drain saving about 15%. While the amount of saving for each unit is small, the aggregate saving is very large, because there are hundreds of such units in each central office. In the over-all redesign, the floor area requirements for the larger offices were reduced by almost 20%.

Transistor Multifrequency Supply

A new transistor multifrequency supply circuit[9] has been standardized for the No. 5 crossbar senders. One supply circuit will be provided in each multifrequency sender. Each supply circuit has six transistor oscillators, one for each of the frequencies used in the multifrequency signalling system. Since transistors require no warm-up time, the oscillators are ready for instant use when the sender is selected. Also, the constant current drain required by vacuum-tube oscillators is not needed. This oscillator unit is omitted from the senders, if the large vacuum-tube type of common supply is required for other equipments in the same building.

New Maintenance Features

To reduce maintenance efforts and to provide better service, the maintenance facilities have been improved, and additions made. Some of the more interesting facilities are described in the following.

Automatic Progression Trunk Test Circuit[10]

Testing of trunk circuits in No. 5 crossbar offices has been accomplished by manual operation of keys to select the desired circuit and to program the test. It was soon evident that a labor-saving automatic test circuit would be economical in the larger offices. Such a test circuit has been designed for control by a perforated teletypewriter tape. For each trunk

Fig. 17. The new dial tone marker

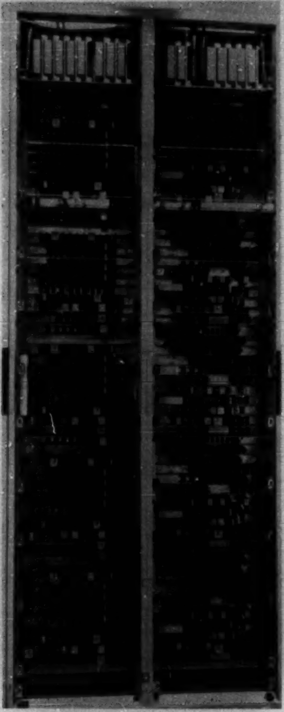


Fig. 18 (above left). The old originating register frame (five registers per frame)

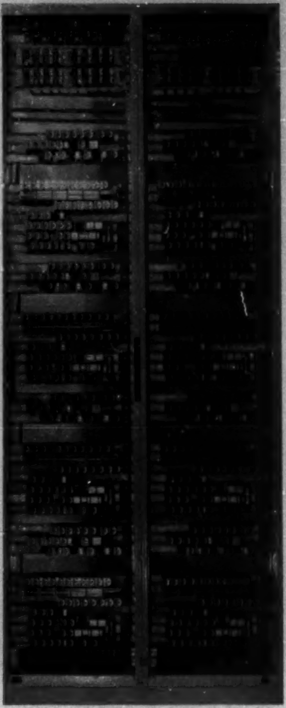
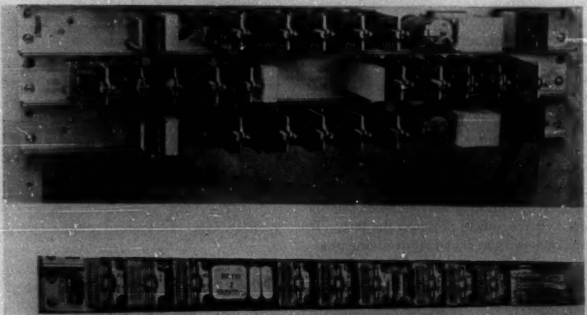


Fig. 19 (above right). The new originating register frame (eight registers per frame)

Fig. 20 (below). Old and new incoming trunk circuit units



circuit to be tested, there are 12 lines of data perforated on the tape. These contain:

1. The location of the trunk circuit on the switches.
2. The program of tests to be applied.
3. Office code and numerical digits for the test call.

The machine reads the data, causes the switching system to connect the test circuit to the trunk circuit, and applies the tests. Failures are reported to the maintenance force by means of visual and audible alarms.

A library of tapes is provided, so that different testing programs may be used on successive days. The tapes are provided in duplicate pairs, so that if one of a pair is damaged or destroyed, its mate will be available for use. The first use, of course, is to perforate automatically another mate tape, so that a pair may be always available.

The original tapes are made at the central office by operating keys on the test frame to represent the data for one trunk circuit. These data are then perforated on the tape. The keys are then set manually for the next trunk circuit, and the tape perforated, etc. This is a slow manual process but is seldom repeated.

With this arrangement the maintenance force can start a test of the trunk circuits which will continue without attention until trouble is encountered. This, of course, saves a substantial amount of maintenance effort in the larger offices.

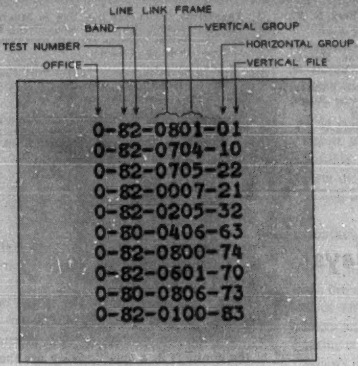
AUTOMATIC LINE INSULATION TEST CIRCUIT[11]

Another interesting maintenance device which has been standardized for No. 5 crossbar (and some other systems) is the automatic line insulation test circuit. This circuit checks the insulation resistance of the wires reaching out to the customer premises, and provides a typed report at the test center of trouble or incipient trouble.

This outside plant of the telephone system deteriorates because of age, weather, and other factors, which causes a reduction in the insulation resistance between

Table I. Dial Tone Marker

	Old Marker	New Marker
No. of 23-inch bays	2	1
Relays	561	456
Machine time	0.43 second	0.26 second
Peak current	28.0 amperes	15.4 amperes
Average current	20.6 amperes	11.6 amperes


LINE LINK FRAME
BAND
VERTICAL GROUP
TEST NUMBER
HORIZONTAL GROUP
OFFICE
VERTICAL FILE

```
0-82-0801-01
0-82-0704-10
0-82-0705-22
0-82-0007-21
0-82-0205-32
0-80-0406-63
0-82-0800-74
0-82-0601-70
0-80-0806-73
0-82-0100-83
```

Fig. 21 (left). Typed record from line insulation test frame

Fig. 22 (right). Intraoffice trunk circuits with ringing selection switches; 40 circuits per bay

the conductors of a pair, and between these conductors and neighboring pairs in a cable, or between the conductors and ground. These conditions may be caused by moisture entering cable sheath breaks, dust on cable terminal face plates, tree branches, dirt on insulators of open-wire lines, and by breaks in the covering of drop wires. In dry weather these defects do not affect service. However, if not corrected, a sudden rainy period will put a number of lines out of service because of low insulation resistance. This is not only bad from a service standpoint, but it may impose a serious overload on the maintenance people, possibly at night, or on a nonworking day.

To find these incipient troubles before they affect service, and so that they can be repaired during regular working hours, the automatic line insulation testing equipment is being installed in most of the new No. 5 crossbar offices, and is being added in many of the existing offices. It can be started either manually or by means of a program clock at the central office, or manually from a remote test bureau. The report of the failures is automatically teletyped to the test bureau, and the details of the test record appear on a printed page, a sample of which is shown in Fig. 21.

The clock is used principally to start a test in the early morning when low temperature has caused the cable to "breathe in" damp air through sheath breaks. This reduces insulation resistance of conductors near the break. The control from the remote test bureau, is used principally to start the test when sudden weather changes affect the outside plant. The rate of testing is 10,000 to 12,000 lines per hour so that all lines in an office can be tested during a short rainy period. The data at the test bureau are analyzed, and further tests and inspections are made of suspected parts of the plant.

The test circuit has a nongrounded d-c source to detect leaks between the conductors of a pair, and a grounded source to find leaks to ground from either conductor. It is also arranged to test for leaks to other conductors in the same cable. These other conductors are usually connected to a negative 48-volt source at the central office; the leaks can therefore be distinguished from leaks to ground. A filter is provided to prevent interference from 25- and 60-cycle earth potentials so that a-c earth potentials will not exaggerate leakage to ground. The circuit is capable of detecting insulation resistance as high as 10 megohms, but is usually used at 2 megohms or less.

Fig. 22 shows intraoffice trunk circuits. Fig. 23 shows the new fuse panel, which will be smaller than its predecessor. It has a dead front, and better appearance; fuses can be replaced without tools, and there can be no flashover between fuses.

Fig. 24 shows a typical crossbar switch. The armatures of the newest crossbar switch have damping cones which deflect the select finger out of its vertical travel path. The finger strikes the edges of the actuating cards and loses its energy quickly. Fifteen milliseconds of marker work time are saved if all switches in the office are provided with damping cones. This is because the marker does not have to wait as long before reoperating the hold magnet which has just released the select finger on a previous call.

In this paper it has been possible to cover only the highlights of the recent changes in the design of the No. 5 crossbar system. In some cases the more interesting or more important items were chosen for discussion. In other cases a few items, which are typical of many others, were described.

References

1. The Number 5 Crossbar Dial Telephone Switching System, F. A. Korn, James G. Ferguson. *AIEE Transactions*, vol. 69, pt. 1, 1950, pp. 244–54.

Fig. 23. The new fuse panel, which will be smaller than its predecessor. Has dead front, better appearance; fuses can be replaced without tools, and there can be no flashover between fuses

Fig. 24. Typical crossbar switch

2. Crossbar Dial Telephone Switching System, **F. J. Scudder, J. N. Reynolds.** *AIEE Transactions (Electrical Engineering)*, vol. 58, May 1939, pp. 179–92.

3. Pretranslation in No. 5 Crossbar, **R. C. Avery.** *Bell Laboratories Record*, New York, N. Y., Apr. 1950, pp. 156–61.

4. Pulse Conversion in No. 5 Crossbar, **H. J. Michael.** *Ibid.*, Dec. 1950, pp. 533–37.

5. Nationwide Numbering Plan, **W. H. Nunn.** *Bell System Technical Journal*, New York, N. Y., vol. 31, Sept. 1952, pp. 851–59.

6. Fundamentals of the Automatic Telephone Message Accounting System, **John Meszar.** *AIEE Transactions*, vol. 69, pt. 1, 1950, pp. 255–69.

7. Foreign Area Translation in No. 5 Crossbar, **C. F. Knepper.** *Bell Laboratories Record*, New York, N. Y., Feb. 1955, pp. 54–57.

8. Code Translation in No. 5 Crossbar, **J. A. Ceonzo.** *Ibid.*, Dec. 1954, pp. 471–74.

9. Transistor Oscillator for Use in Multifrequency Pulsing Current Supply, **F. H. Blount.** *Bell System Technical Journal*, New York, N. Y., Nov. 1953, pp. 1313–31.

10. Automatic Progression Trunk Test Circuit, **H. D. Cahill.** *Bell Laboratories Record*, New York, N. Y., Aug. 1954, pp. 313–18.

11. Automatic Line Insulation Test Circuit, **J. W. Dehn, R. W. Burns.** *Bell System Technical Journal*, New York, N. Y., May 1953, pp. 627–46.

Capability of Sealed Contact Relays

O. M. HOVGAARD
NONMEMBER AIEE

A CIRCUIT designer contemplating the use of a relay must consider its operating characteristics and the reliability of its contacting function. The electric erosion causing contact wear gives rise to contact surface conditions permitting contact failure to open a circuit caused by locking; particularly in talking circuits, reliability can be further impaired by the operating environment which may contain corrosive elements, organic vapors, or mobile dirt, all of which increase the probability of failure to give contact closure. Unfortunately, the incidence of these factors affecting reliability is not predictable with respect to time or place. The considerations involved are not unlike those associated with electron tubes and transistors. Because of the growth in the complexity of electric switching systems increasing emphasis is placed upon the elimination of factors affecting component reliability. In the Bell System one means to achieve improved reliability has been the development of sealed contacts. Two types have reached application, both providing encapsulation to free contacts from the effects of environment; one type, in addition, circumvents the effects of electric erosion by employing liquid contact surfaces. The development of sealed contacts and relays using such contacts has been described elsewhere;[1–4] the purpose of this paper is to consider some of their characteristics from the standpoint of application.

Fig. 1 shows three forms which these contacts have assumed. *A* is the simple encapsulation of a more or less conventional contact; *B* and *C* are forms employing mercury-wetted contact surfaces. The basic principle of operation of all three designs is that of overlapping cantilevers of magnetic material separated by a small gap. When the contact is surrounded by an operating winding carrying current, the ends of the cantilevers are drawn together to form the contact pair controlling a load circuit.

Operating Characteristics

For the type of sealed contact under consideration, the operating characteristics are defined in terms of its behavior when it is associated with a simple air-core coil of specified dimensions. The characteristics are expressed in terms of that value of ampere-turns at which the contact will just close and that lower value at which it will just reopen.

In the course of development, the designer will have established the desired mean values of these characteristics. Through a visualization of the process to be used in manufacture and by laboratory experiments, he will also have determined the attainable standard deviations for these characteristics and a measure of the closeness within which the mean values can be held. Upon these considerations he will have specified the desired mean values and the limits with respect to these within which the manufactured product must be held.

Relay Design

While for manufacturing purposes the characteristics of the sealed contact are defined in terms of the contact behavior in an air-core coil, actual relay designs usually provide ferromagnetic paths to increase electromechanical efficiency and to decrease magnetic coupling to nearby circuit components. More than one contact may be used within the same operating coil. The operating characteristics which can be ascribed to a specific design will depend upon the number of contacts, the characteristics of the contacts and their relative positions within the coil, the reproducibility of the magnetic circuit and coil turns, etc.

For a given relay design, the spread in the sealed contact operating characteristics and the combination and placement of the contacts in a specific relay are the main factors causing a spread in the relay operating characteristics about their mean values. Laboratory studies make it possible to estimate the magnitude of these variables and to arrive at tolerances which the relays must meet in manufacture.

What should the relay tolerances mean to the circuit designer and how may he use them? Since sealed contacts are not adjustable after manufacture, the answer depends on how the manufacturer has used the limits given in the specification.

Suppose that the upper limit of a relay design requires a current I to give contact closure, and that there is no information concerning the distribution of this relay operating characteristic. The relay is to be used in a circuit subject to voltage and temperature variations. Let it be assumed that nothing is known about the distribution of these variables with time but that their maximum values can be stated. The coincidence of the highest temperature condition and lowest voltage supply, barring other variables, is called the worst circuit condition since it results in the minimum value of current for the operation of the relay. How shall this worst circuit current be related to the value I defined in the foregoing?

Lacking information to the contrary about the relay characteristic, the circuit designer can only assume that in manufacture the upper value of required current I is as permissible as any other value within the stated tolerance and that he

---

Paper 56-658, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 28, 1956; made available for printing April 23, 1956.

O. M. Hovgaard is with the Bell Telephone Laboratories, Inc., New York, N. Y.

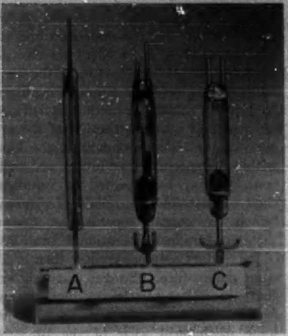


Fig. 1. Typical sealed contacts

will have to provide a margin between it and the worst circuit current. The size of the margin represents the designer's judgment of the reliability of the facts available to him and maybe also his estimate of the probability of occurrence of the worst circuit current.

However, if the designer knows that the manufacturer has designed his process so that the specified quality characteristic meets the standard deviation and maintenance of mean value visualized in the specification and if he can be assured that the manufacturer maintains the characteristic in statistical control at the desired level, he is then entitled to inference with respect to the distribution of the quality characteristic within the stated limits. For purposes of illustration, assume that the distribution of the characteristic is centered and can be represented by a normal distribution and that the manufacturing limits represent four standard deviations. If then the worst circuit current were placed at the upper limit for the relay I, the probability of a failure under worst circuit conditions would be about one in 30,000. This, combined with the likelihood of occurrence of the worst circuit condition, might be more conservative than is required. Circuit power could be conserved by placing the worst circuit condition at, say, $2^{1}/_{3}$ standard deviations from the central value of the operate characteristic and then the probability of relay failure under worst circuit conditions would be about one chance in a hundred. If this risk is acceptable, it becomes possible to approve a circuit design which might not be possible if there were not assurance that the relay characteristic had a normal distribution.

As a goal, if the circuit variables and component characteristics are in fact in statistical control at the desired levels, the designer can determine circuit performance and reliability on a statistical basis. This will permit him to approve smaller circuit margins than would otherwise be possible, thereby obtaining more complete utilization of component capability as well as improved circuit efficiency.

Contact Behavior

The operating characteristics discussed are the result of design and process decisions concerning dimensions and tolerances and, thus, are subject to manufacturing control. The behavior of the contact with respect to its electric load as well as its reliability in use result from design decisions concerning contact materials, purity and kind of selected gaseous environment, contact surface preparation, cleanliness, etc. The relationships between these factors and the contact life and reliability are not as well understood as are those affecting the operating characteristics and, therefore, life and reliability are not yet subject to manufacturing control to the same degree as are the operating characteristics.

Consider first the behavior of type-*A* contacts. Life test studies with electric loads of various magnitude and kind show that, like other contacts having solid surfaces, they suffer erosion which becomes more violent as the loads increase. This erosion progressively alters the surface geometry causing build-ups on one contact and corresponding pits on the other. Unlike ordinary contacts the by-products of erosion are magnetic and tend to collect in the gap under the influence of the operating flux. As a result, there are at least two mechanisms which can cause contact failure. In one, build-ups and pits may reach the right configuration to cause mechanical locking which, if it cannot be overcome by the available reed-restoring forces, prevents opening of the contact. In the other, the by-products of erosion, which may be either loose or fastened to one reed, can form chains of particles capable of bridging the space between the unoperated reeds which then will be electrically closed. As a failure mechanism, locks are several times as numerous as bridges, but the relative expectation of their occurrence appears unrelated to the length of time that the contact has been in use.

In general, the behavior with load is quite similar to that observed with other hard-surface contacts. As the load increases, the life of the contact gets shorter. However, for the upper load level of which these contacts are capable, the rate of decrease of life with increasing load is not as great as that observed with ordinary contacts. This may be because of the capture of magnetic erosion by-products in the operating gap. For part of the operations, contact will be established through such particles and, as a result,

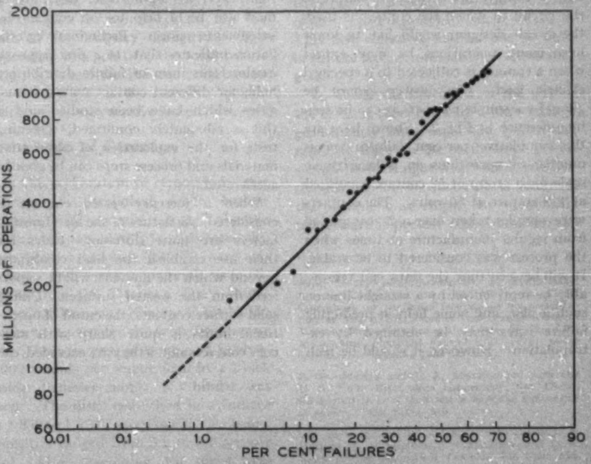


Fig. 2. Life study of type-A contacts

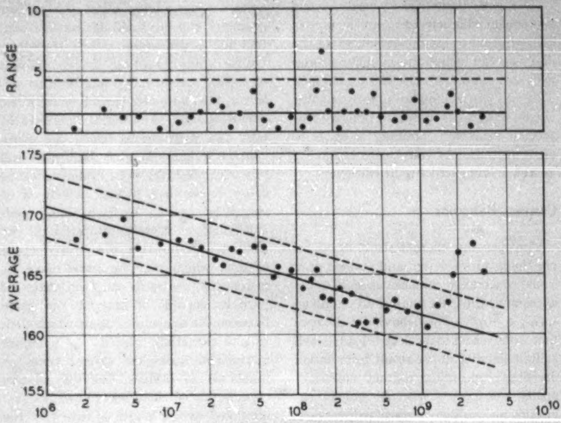


Fig. 3. Stability of type-B contact with life

they may absorb a part of the erosion. These particles also, in general, tend to move at each operation and, by spreading the erosion over a greater portion of the contacting area, may effectively increase the life of the contact.

As might be expected from the complicated behavior described, observations of laboratory life tests show an enormous spread in the life obtained from nominally identical sealed contacts. Since the end of contact life means the failure of the circuit in which the contact is used, the circuit designer would like to know how many operations he may expect when a contact is subjected to a specified electric load. The answer cannot be given by a simple number, as can be seen by reference to Fig. 2. Shown here are the cumulative per-cent failures versus number of operations on a logarithmic scale for a group of 52 contacts operated at 1/8 ampere at 50 volts. The contacts were samples taken over a 2-year period from regular manufacture at times when the process was considered to be stable. It will be seen that the data can reasonably be represented by a straight line on such a plot, and some help in predicting failure rates may be obtained by extrapolation. However, it should be realized that this technique involves some risks. In the first place, the confidence limits are rather wide with so small a sample and, secondly, another sample might yield a different slope. Also there can be no assurance that there is not some other mechanism which could cause early failures in excess of those predicted. It may not have been possible to detect such a cause in so small a sample. The linear relationship, which represents a log-normal distribution, may also be of help to the contact development engineer. Preliminary investigation indicates that, to a first approximation, this form of failure distribution holds for different contact materials and gases which have been studied and if this is adequately confirmed, screening tests for the exploration of alternative materials and process steps can be greatly accelerated.

When a mercury-wetted contact is considered, the nature of the load limiting factors are quite different. Rules for their use establish the load conditions beyond which the mercury will be vaporized from the wetted surfaces. Unlike solid-surface contacts, the cutoff of operational limits is quite sharp with mercury contacts and, if they are exceeded, the contact life is drastically reduced. However, within the recommended operating range, there is in effect no electric erosion since, at each operation, the mercury film is restored through capillary feed paths under the action of surface tension. As a result, the life of a mercury contact is independent of the load within the specified load limits, and such contacts may be expected to maintain their contact function reliably for many billions of contact operations. Under these conditions, operational stability of the contact assumes greater than ordinary importance. Studies have shown that, in spite of the cushioning effects of liquid contact surfaces, there is some mechanical wear of the surfaces supporting the mercury film, and the pretensioned armature spring may lose some of its initial force because of fatigue. These factors cause slow systematic changes in the operating characteristics. Fig. 3 shows a control chart plot of the operate value of two mercury-wetted contacts.[5] These charts are for the successive averages and range of the two contacts. It is seen that, after 2 billion operations, the characteristics studied went out of control. Although after 3 billion operations the observed operate values were still not greatly different from their initial values, and the contacting function was unimpaired, it was no longer possible to predict operational performance, and on this basis, the useful life must be considered ended. This illustrates the application of a quality control chart as a means for defining life.

In summary, experience with sealed contact relays indicates that statistical methods are useful not only during development and manufacture but also promise to be an effective means for describing their attributes.

References

1. Glass-Enclosed Reed Relay, W. B. Ellwood. *Electrical Engineering*, vol. 66, 1947, pp. 1104–05.

2. Mercury Contact Relays, J. T. L. Brown, C. E. Pollard. *Ibid.*, pp. 1106–09.

3. Balanced Polar Mercury Relay, J. T. L. Brown, C. E. Pollard. *Bell System Technical Journal*, New York, N. Y., vol. 32, no. 6, 1953, pp. 1393–1411.

4. Development of Reed Switches and Relays, O. M. Hovgaard, G. E. Perreault. *Ibid.*, vol. 34, no. 2, 1955, pp. 309–32.

5. Statistical Methods in the Development of Apparatus Life Quality, Enoch B. Ferrell. *AIEE Transactions*, vol. 64, Oct. 1945, pp. 700–03.

Vol.
Pt. 1
1956

Investigations Concerning Polarization in Barium Titanate Ceramics

G. W. MARKS
NONMEMBER AIEE

D. L. WAIDELICH
FELLOW AIEE

L. A. MONSON
NONMEMBER AIEE

Synopsis: Considerable information may be obtained concerning polarization in barium titanate ceramics by a careful analysis of hysteresis loops. Mathematically it is shown that both symmetrical and unsymmetrical loops, the latter being characteristic of polarized specimens, can be represented best by a Fourier series. It was found that direct voltage and even harmonic components are present in the unsymmetrical loops. The per-cent second harmonic and the electromechanical coupling coefficient both rise with increasing depolarizing voltage and reach a maximum value at approximately 15,000 volts per cm (centimeter). The measurement of the second harmonic is possible even when the activity is so small that measurement of the coupling coefficient becomes very inaccurate or impossible. The addition of calcium titanate in small percentages markedly raises the coercive force but lowers the spontaneous and remanent polarization in the range from 0 to 80 C (degrees centigrade). However, it does not have marked effect between the latter temperature and the Curie point. It was found from studies of the discharge currents on depolarizing specimens by sudden immersion in hot oil that the polarization observed from the discharge curves is essentially the same as the remanent polarization determined from oscillograms of saturated hysteresis loops.

IN RECENT years, many investigations have been made in attempts to elucidate the polarization phenomena of ferroelectrics because of their engineering applications. Of particular interest have been those concerning barium titanate and related compounds having the perovskite structure. The polarization of polydomain ceramic ferroelectrics by the application of an external field consists, in general, of the following processes: 1. the direction of polarization in any crystallite tends to orient in the direction of the applied field, 2. the polarization in the domains increases in magnitude, and 3. the relative volume occupied by different domains changes primarily by growth in the forward direction; sidewise motion of 180 degrees side walls hardly ever occurs.[1,2] This includes the growth of the induced polarization, much the same linear process as in ordinary dielectrics.

Among the lag and fatigue effects which characterize ferroelectrics is the hysteresis loop. Much knowledge concerning polarization and related properties may be obtained from a careful study of these loops as recorded by means of a cathode-ray oscillograph.[3–12] However, the rate of change of the field for a 60 cycle-per-second hysteresis loop is so rapid that long-time relaxation phenomena which contribute to the polarization are not observed and thus somewhat lower polarization values than the possible maxima are obtained. Recently, by the use of a high-voltage switching and integrating circuit, the relationship between total and remanent polarization to the applied field and applied field duration has been determined at constant temperature for barium titanate ceramics.[13]

For ready reference a representative hysteresis loop for a polycrystalline multidomain ferroelectric is shown in Fig. 1. The variation of polarization P with field strength E is depicted. The spontaneous polarization P_s is obtained by extrapolation of the saturated linear portion of the loop to zero field strength. Note that beyond E, where the linearity commences, the increase in polarization is induced polarization only. Except for the fact that this extrapolated portion does not pass through the origin, it is the same as for ordinary dielectrics. Corresponding to a point E_2, the total polarization is P_t and the induced polarization $P_i = P_t - P_s$. P_r is the remanent polarization; E_c is the coercive force.

Preparation of Ceramics

In the preparation of ceramics for this investigation, technical grade barium titanate, 97% (per cent); technical grade calcium titanate (96.5%); and chemically pure barium titanate (99.7%) were employed (All ceramic materials were obtained from the Titanium Alloy Manufacturing Company, Niagara Falls, N. Y.) Disks were prepared for firing by pressing 6 to 10 gm (grams) of sample at 5,000 to 8,000 pounds per square inch in a 1-inch inside diameter mold. No binder was used. The disks were fired in a furnace at 1,350 to 1,370 C from one to two hours. Densities ranged from 5.3 to 5.6 gm per cm^3. After machining, the disks were foiled with silver.

Circuit for Recording Hysteresis Loops

For the study of hysteresis loops, in general, previous workers have used a parallel-connected resistance-capacitance voltage divider. However, it was found that a circuit utilizing capacitance voltage dividers only was more suitable for the present purpose. A diagram of this circuit is shown in Fig. 2. Maximum secondary peak voltages of approximately 25 kv per cm could be impressed. Secondary voltages were controlled by means of a Variac. The capacitance voltage divider, consisting of C_1 and C_2, resulted in a voltage across C_2 directly proportional to the alternating voltage across the ceramic and this voltage was applied to the horizontal plates of the oscillograph. This voltage is proportional to the electric field E applied to the ceramic. The capacitance of C_2 being much larger than that of C_1, its effect on the voltage across C_1 may be neglected. Thus, the field intensity across the ceramic is $E = C_1V_2/C_1d$. V_2 is the voltage across C_2 and d is the thickness of the ceramic. C_3 integrates the instantaneous polarizing current i which flows into the ceramic to give a voltage V_3 which is proportional to the polarization P. The capacitance C_4 of the ferroelectric is extremely high compared to that of C_3. V_3 is then the voltage applied to the vertical plates of the oscilloscope and so the hysteresis loop is observed.

If P is the polarization in coulombs per cm^2 and a is the area in cm^2 of an electrode face, then $i = a(dP/dt)$ and $P = (1/a)\int idt$. Since $V_3 = 1/C_3 \int idt$, $P = C_3V_3/a$. This last equation enables one to calibrate the axis of ordinates of an oscillogram.

To check the operation of the circuit the ceramic is replaced by a high-voltage good-quality mica capacitor. Then the resulting trace on the oscilloscope screen should be a straight line with a positive slope when the vertical gain is zero. If the pattern is an ellipse, an undesirable phase shift is present somewhere in the

---

Paper 56-682, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted February 21, 1956; made available for printing June 12, 1956.

G. W. MARKS and L. A. MONSON are with the U. S. Navy Electronics Laboratory, San Diego, Calif., and D. L. WAIDELICH is with the University of Missouri, Columbia, Mo.

The authors take the opportunity to thank J. F. Haskins, T. M. Lambert, J. S. Hickman, and D. E. Risty of the Acoustical Division for their aid, suggestions, and encouragement during the course of this work.

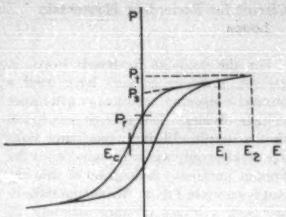


Fig. 1. Representative hysteresis loop for unpolarized ceramic showing variation of polarization P with field strength E. Spontaneous polarization is P_s, remanent polarization P_r, and coercive force E_c

circuit or in the oscilloscope, and this should be eliminated.

Application of Fourier Series

With varying degrees of success, attempts have been made by a number of investigators to fit various mathematical equations to the magnetization curve of ferromagnetics. Press[11] and Lewis[13] have suggested the use of Fourier series, and in the analysis of the ferroelectric hysteresis loops of barium titanate such a procedure has been used. Accordingly, if

$$x=\left(\frac{E}{E_m}\right)\sin \omega t \qquad (1)$$

where E_m is the maximum electric field applied to the specimen and ω is the angular frequency of the sinusoidal variation producing the loop, then

$$y=\frac{P}{P_m}=a_0+a_1\cos \omega t+a_2\cos 2\omega t+\ldots+ b_1\sin \omega t+b_2\sin 2\omega t+\ldots \qquad (2)$$

where P_m is the maximum polarization and the a's and b's are constants for a given hysteresis loop. (It is perhaps well to point out that for barium titanate the dielectric constant is so large that the displacement $D\approx 4\pi P$ in centimeter-gram-second units. D and P can be used interchangeably if meter-kilogram-second units are used.) If the parameter t is eliminated between equations 1 and 2

$$y=(A_0+A_1x+A_2x^2+A_3x^3+\ldots)\pm (1-x^2)^{\frac{1}{2}}(B_0+B_1x+B_2x^2+ B_3x^3+\ldots) \qquad (3)$$

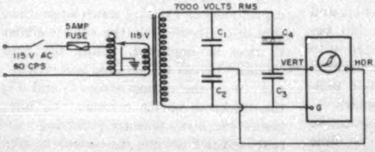


Fig. 2. Circuit for studying hysteresis loop of ferroelectrics. C_1 0.003 μf, 25 kv; C_2 2.0μf, 600 volts; C_3 7.2 μf, 600 volts; C_4 ferroelectric

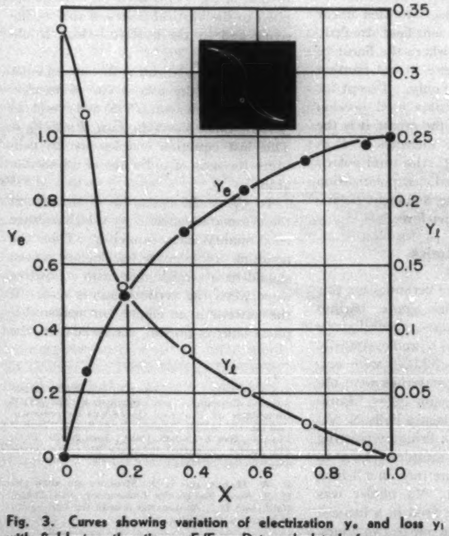


Fig. 3. Curves showing variation of electrization y_e and loss y_l with field strength ratio $x=E/E_m$. Data calculated from measurements on hysteresis loop (insert) for chemically pure ceramic disk of unpolarized BaTiO₃. E_m=11,980 volts per cm. Calibration cross not shown

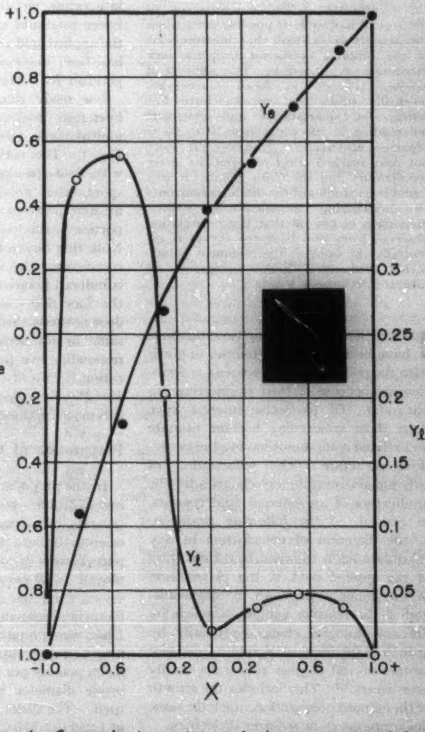


Fig. 4. Curves showing variation of electrization y_e and loss y_l with field strength ration $x=E/E_m$. Data calculated from measurements on hysteresis loop (insert) for technical grade ceramic disk of polarized BaTiO₃. E_m=9700 volts per cm. Calibration cross not shown

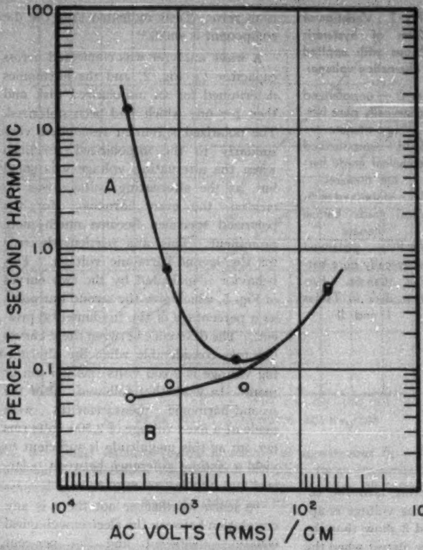


Fig. 5. Per-cent second harmonic (percentage of the fundamental voltage present) as function of voltage applied to ceramic for: A—polarized, and B—unpolarized barium titanate disks

where

$A_0 = a_0 + a_2 + a_4 + a_6 + \ldots$
$A_1 = b_1 + 3b_3 + 5b_5 + 7b_7 + \ldots$
$A_2 = -2a_2 - 8a_4 - 18a_6 - 32a_8 - \ldots$
$A_3 = 4b_3 - 20b_5 - \ldots$, etc.

The ± signs in equation 3 correspond to the upper and lower parts of a given hysteresis loop respectively. It is possible to eliminate the ± signs entirely by squaring equation 3 to obtain

$$y^2 - 2y(A_0 + A_1x + A_2x^2 + \ldots) + (D_0 + D_1x + D_2x^2 + D_3x^3 + \ldots) = 0 \qquad (4)$$

in which

$$\Sigma D_n x^n \equiv (\Sigma A_m x^m)^2 - (1 - x^2)(\Sigma B_m x^m)^2$$

Equation 4 is the equation of the entire hysteresis loop but for most purposes equation 3 is more useful.

A symmetrical hysteresis loop is one in which the origin of co-ordinates is a center of symmetry and so involves only odd functions. When this fact is utilized, the even A's and odd B's are zero so that

$$y = (A_1x + A_3x^3 + A_5x^5 + \ldots) \pm (1 - x^2)^{\frac{1}{2}}(B_0 + B_2x^2 + B_4x^4 + \ldots) \qquad (5)$$

Let y_a represent the upper portion and y_b the lower portion of the symmetrical loop, then

$$(A_1x + A_3x^3 + \ldots) = \tfrac{1}{2}(y_a + y_b) = y_e \qquad (6)$$

and

$$(1 - x^2)^{\frac{1}{2}}(B_0 + B_2x^2 + \ldots) = \tfrac{1}{2}(y_a - y_b) = y_l \qquad (7)$$

where y_e represents the electrization curve (similar to the magnetization curve of ferromagnetism) and y_l represents the loss curve corresponding to the loss as shown by the area of the hysteresis loop.

As an illustration of the procedure to be employed, measurements were first made on the symmetrical hysteresis loop shown in the insert of Fig. 3. Values of y_e and y_l were then calculated and the resulting electrization and loss curves are shown in this figure. The A's of equation 6 and the B's of equation 7 may be calculated by selecting a number of points on the curves of Fig. 3, and using the number of A's and B's equal to the number of points selected. In most cases, however, the experimental curves for y_e and y_l may be fitted adequately with the use of a lesser number of constants by employing special functions such as the hyperbolic tangent or exponential functions.

Polarized specimens of ceramic barium titanate produce unsymmetrical hysteresis loops if the applied alternating voltage is sufficiently high. For an unsymmetrical loop

$$y_a = (A_0 + A_1x + A_2x^2 + \ldots) + (1 - x^2)^{\frac{1}{2}} \times (B_0 + B_1x + B_2x^2 + \ldots) \qquad (8)$$

$$y_b = (A_0 + A_1x + A_2x^2 + \ldots) - (1 - x^2)^{\frac{1}{2}} \times (B_0 + B_1x + B_2x^2 + \ldots) \qquad (9)$$

Then

$$y_e = \frac{1}{2}(y_a + y_b) = (A_0 + A_1x + A_2x^2 + \ldots) \qquad (10)$$

and

$$y_l = \frac{1}{2}(y_a - y_b) = (1 - x^2)^{\frac{1}{2}} \times (B_0 + B_1x + B_2x^2 + \ldots) \qquad (11)$$

Electrization and loss curves derived from the hysteresis loop of a polarized disk of barium titanate are shown in the insert of Fig. 4. The upper part of the hysteresis loop is the saturated part, i.e., the direction of polarization and the direction of the applied a-c field are then the same. The larger lower part of the hysteresis loop occurs when the a-c voltage is opposite in direction to that of polarization

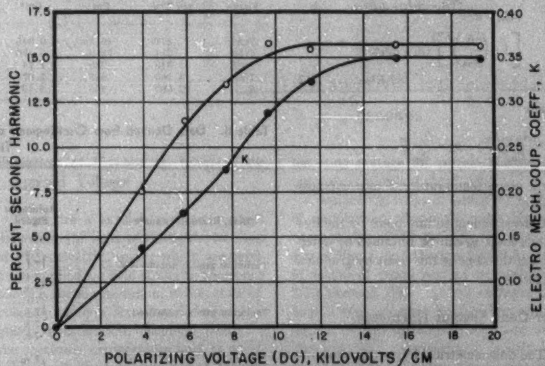


Fig. 6. Rise of percentage second harmonic present and of electromechanical coupling coefficient with increase in polarizing voltage for barium titanate disk

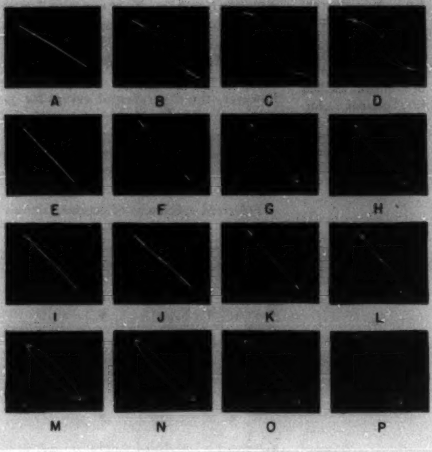


Fig. 7. Variation of form of hysteresis loops with applied alternating voltage:

A–D — unpolarized chemically pure barium titanate
E–H — unpolarized technical grade barium titanate
I–L—polarized technical grade barium titanate
M–P — polarized chemically pure barium titanate. Further data in Tables I and II

The *A*'s of equation 10 and the *B*'s of equation 11 may be determined by selecting a sufficient number of points from Fig. 4. The unsymmetrical shapes of the y_c and y_t curves show that the differential permittivity is larger in the unsaturated direction than it is in the polarized direction.

For theoretical work equation 6 is very useful. Once the constants for any given loop have been determined it is possible to obtain the differential susceptibility, $\eta_d = \partial P/\partial E$. From equation 3 it follows that

$$\left(\frac{\partial P}{\partial E}\right)_{T,C} = \alpha_0 + \alpha_1 E + \alpha_2 E^2 + \alpha_3 E^3 + \ldots \pm \frac{E}{(E_m - E^2)^{1/2}} (\beta_0 + \beta_1 E + \beta_2 E^2 + \ldots) \pm \left[1 - \left(\frac{E}{E_m}\right)^2\right]^{1/2} (\beta_1 + 2\beta_2 E + 3\beta_3 E^2 + \ldots) \quad (12)$$

where

$$\alpha_n = \frac{A_{n+1} P_m}{E_m^{\,n+1}}; \quad \beta_n = \frac{B_n P_m}{E_m^{\,n}}$$

for a given temperature *T* and composition *C*.

Corresponding to any point on the hysteresis loop, η_d can be obtained by determining the slope at this point by graphical methods.

Per-Cent Second Harmonic

The unsymmetrical hysteresis loops observed for polarized barium titanate suggest that direct and even harmonic voltage components may be present as a result of the charge flowing into the ceramic when an alternating voltage is applied. Equations 3 and 5 show that the even *A*'s and odd *B*'s are absent when the loop is symmetrical but are present when the loop is asymmetrical. Since the even *A*'s and odd *B*'s are sums of the amplitudes of the direct and even harmonic components, it would seem that some of these components should then be different from zero. Tests indicated that the d-c component is small.[16]

A wave analyzer was connected across capacitor C_3, Fig. 2, and the harmonics determined for an unpolarized disk and then for one which had been polarized. The polarized specimen responded very similarly to the unpolarized specimen when the alternating voltage was small but, as the alternating voltage was increased, the even harmonics for the polarized specimen became much more prominent. This was particularly true for the second-harmonic voltage. This behavior is indicated by the two curves of Fig. 5, which give the second harmonic as a percentage of the fundamental present. The difference between these curves becomes considerable when the alternating voltage is 2,000 volts rms per cm or more. In work that followed, all of the second-harmonic measurements were made at a fixed voltage of 2,500 volts rms per cm as this magnitude is sufficient to yield a distinct difference between polarized and unpolarized specimens.

To learn whether or not there is any correlation between the electromechanical coupling coefficient and the per-cent second harmonic, an initially unpolarized specimen was polarized at successively increasing direct voltages for one hour at room temperature.

After each period of polarization, the per-cent second harmonic present was

Table I. Data Derived from Oscillograms of Hysteresis Loops for Disk of Unpolarized Chemically Pure Ceramic Barium Titanate

Reference Figure	Maximum Field Strength, E_m, Volts per Cm	Coercive Force, E_c, Volts per Cm	Maximum Polarization P_m, Coulombs per Cm²	Remanent Polarization P_r, Coulombs per Cm²	Average Dielectric Constant, P_m/E_m	Center Dielectric Constant	End Dielectric Constant
7(A)	318	10.2	0.045	0.001	1,600	1,530	1,600
7(B)	714	73.6	0.170	0.021	2,700	4,330	1,450
7(C)	1,610	286	1.74	0.521	12,200	23,900	1,730
7(D)	4,200	525	5.07	1.55	13,600	41,200	2,380
3	12,000	955	8.42	2.8	7,900	45,300	1,670

Table II. Data Derived from Oscillograms of Hysteresis Loops for Disks of Ceramic Barium Titanate

Disk, Barium Titanate	Reference Figure	Maximum Field Strength E_m, Volts per Cm	Maximum Polarization Coulombs per Cm²	Average Dielectric Constant
Technical grade, unpolarized	7-e	339	0.046	1,520
	7-f	953	0.149	1,770
	7-g	2,940	0.638	2,450
	7-h	6,330	2.03	3,620
Technical grade, polarized	7-i	347	0.054	1,750
	7-j	773	0.126	1,850
	7-k	1,930	0.394	2,320
	7-l	3,960	1.13	3,250
		5,200	2.35	5,200
Chemically pure, polarized	7-m	559	0.128	2,580
	7-n	1,770	0.544	3,480
	7-o	4,080	2.016	5,590
	7-p	9,360	5.54	6,690
		14,700	12.0	9,240

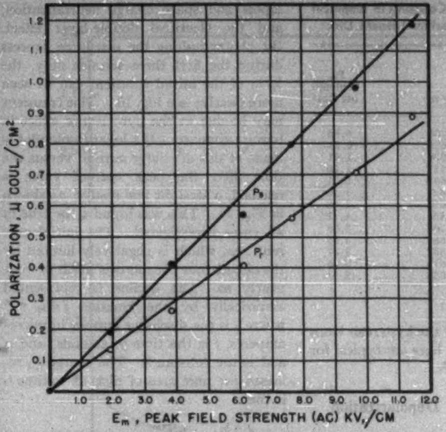


Fig. 8. Linear variation (below saturation) of spontaneous P_s and remanent P_r polarizations with peak field strength E_m for an unpolarized technical grade barium titanate disk. (Data for specimen other than those of Tables I and II)

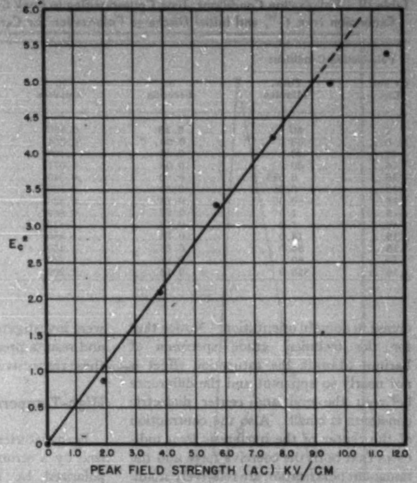


Fig. 9. Linear variation (below saturation) of square of coercive force E_c^2 with peak field strength for unpolarized technical grade barium titanate disk. (Data for specimen other than those of Tables I and II)

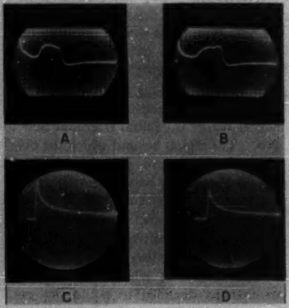


Fig. 10. Representative oscillograms showing depolarization of barium titanate on immersion in oil at 150 C: A, B—0 to 30 seconds; C, D—0 to 3.5 seconds

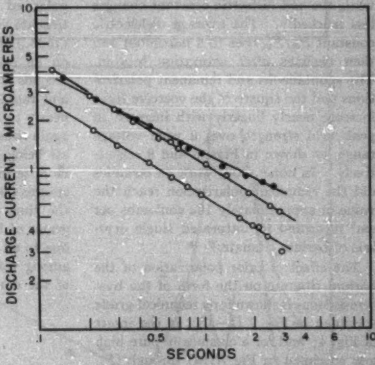


Fig. 11. Logarithmic plot of discharge current versus time after peak current value is reached and before major depolarization of disk occurs

measured and the electromechanical coupling coefficient calculated from measured values of the resonant and antiresonant frequencies for the radial mode of vibration. Results are presented in Fig. 6 for the particular experimental conditions described. Both variables rise with increasing polarizing voltage and reach a maximum value at about 15×10^3 volts per cm. Above this voltage no significant improvement in polarization occurs. The per-cent second harmonic rises somewhat faster than the electromechanical coupling coefficient.

Variation of Hysteresis Loops With Applied Voltage

There are a number of factors which determine the form of the hysteresis loop of a ferroelectric, among which are maximum applied a-c field strength, frequency above a certain minimum value, state of polarization if a ferroelectric ceramic is under investigation, chemical purity of the specimen, temperature and, to a certain extent, the total number of domains present. The variation in the form of observed hysteresis loops with applied alternating voltage for certain barium titanate ceramics is shown in Fig. 7.

Consider the loops obtained for a disk of unpolarized, chemically pure barium titanate, Fig. 7(A) through (D) and insert of Fig. 3, and those obtained for one of technical grade barium titanate Fig. 7(E) through (H). At low voltages the loop is very nearly a straight line and is similar to that for an ordinary dielectric. The resultant polarization is largely induced and the domain structure is but little affected. As the voltage increases, the center of the loop expands with in-

Table III. Polarization Conditions, Time Corresponding to Peak Current, Constants in Empirical Expression $i=i_0\, t^{-m}$, and Initial Discharge Polarization for Ceramic Barium Titanate Disk

Polarization Conditions					
Kv per Cm	Time, Minutes	t_1, Seconds	i_0, Amperes	$-m$	P_x, Coulomb per Cm²
4	60	0.10	0.670	0.735	0.66
8	60	0.05	0.654	0.626	0.73
12	60	0.08	1.176	0.626	1.91
16	60	0.08	1.071	0.627	1.29
16	0.083	0.10	0.343	0.639	0.21
16	0.25	0.10	0.869	0.621	1.13
16	0.50	0.14	0.869	0.621	1.08
16	1.0	0.12	0.869	0.821	1.11
16	5.0	0.12	1.412	0.565	3.06
16	15.0	0.16	1.450	0.557	3.28
16	30.0	0.16	1.450	0.557	3.37
16	60	0.14	1.275	0.615	2.45
16	120.0	0.14	1.209	0.553	2.17

crease in domain orientation. Notice that for the technical grade specimen of barium titanate the saturation effect is not nearly so apparent and the difference between the end and center dielectric constants is small. Also the contraction of the center of the hysteresis loop indicates that both the coercive force and the remanent polarization are relatively small. Data derived from these oscillograms are given in Tables I and II respectively. Note that the center dielectric constant, Table I, increases to very large values while the end dielectric constant changes less markedly. The average dielelectric constant P_m/E_m rises to a maximum and then declines after saturation sets in. The spontaneous and remanent polarizations and the square of the coercive force increase nearly linearly with increase in peak field strength over a wide voltage range, as shown in Figs. 8 and 9 respectively. In none of our work on ceramics did the remanent polarization reach the value of approximately 16μ coulombs per cm² measured for saturated single crystals of barium titanate.[17, 18]

The effect of prior polarization of the barium titanate on the form of the hysteresis loops is shown for a technical-grade specimen in Fig. 7 (I–L); in the insert of Fig. 4; and for a chemically pure high loss specimen in Fig. 7(M) through (P). The noticeable asymmetry in the loop does not appear until the effect of the domain orientation becomes marked which would be in the neighborhood of 3,500 peak volts per cm. For the high-loss specimen the minor axis of the ellipse has expanded considerably over that for low-loss specimens. An ellipse formed at low voltages indicates that the specimen has a relatively high loss.[4] Variation of maximum polarization and average dielectric constant with field strength are given in Table II for these last two disks.

Several specimens of chemically pure and of technical grade barium titanate were investigated and the hysteresis loops and results presented here are typical for these respective grades.

High-Temperature Depolarization

Ferroelectrics which have been polarized by a strong d-c field are readily depolarized by heating at temperatures above the Curie point. We have made certain observations concerning this depolarization process. The experimental procedure was to immerse a previously polarized specimen of ceramic barium titanate very suddenly into hot castor oil (150 C) and record the discharge current on an oscillogram. During the first 3 seconds after immersion there is a sudden and rapid loss of charge which declines nearly to zero and thereafter the current again rises and declines, the second process being much slower and taking 10 or more seconds; see Fig. 10. In some instances the discharge curve dropped below the time axis, as shown in the oscillograms of Fig. 10. The second process is due to all or any of a number of factors among which are the slow disorientation of the domain structure of the ceramic, dipole and space charge neutralization, and the electrical double-layer effect. By photographing the discharge process during the first three seconds only, the form of the initial transient can be seen more clearly; see Fig. 10. This transient may be due to the pyroelectric effect in barium titanate. If a logarithmic plot is made of this discharge current versus the time after the peak current value is reached, a straight line results, as shown in Fig. 11. This was found to be true in all cases investigated. The initial current rise, which is negatively limited by the circuit time constants is linear or very nearly so. The decline is represented empirically by the function $i=i_0\ t^{-m}$ where i is the discharge current in microamperes, t is the time in seconds, and i_0 and m are constants. The charge P_x released per unit area of disk up to time t_2 is then

$$P_x=\left(\frac{1}{2a}\right)i_1t_1+\frac{i_0}{a}\int_{t_1}^{t_2} t^{-m}dt$$

$$=\left(\frac{1}{2a}\right)i_1t_1+\frac{i_0}{a(1-m)}\,[t_2^{1-m}-t_1^{1-m}] \quad (13)$$

The time at which the current reaches its peak is t_1, and the area of one face of the disk is a. The charge resulting from the initial rise is determined from the area of the underlying triangle.

In the oscillograms of Fig. 10 it is observed that the second discharge commences before the initial transient falls to zero. The region beyond which the overlapping of the two processes becomes serious is perhaps best indicated by the fact that the logarithmic plot is then no longer linear. Thus, where P_x is to be determined for purposes of comparison the upper time limit t_2 has been taken to be that value at which the discharge current has declined to 0.1μa per cm.² Some comparative results are presented

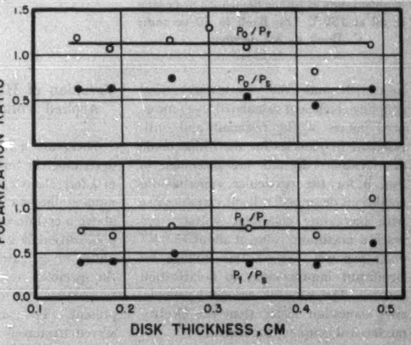


Fig. 12. Polarization ratios versus disk thickness for technical grade of barium titanate ceramics. P_0, P_t = positive and total polarizations from discharge oscillograms; P_r, P_s = remanent and spontaneous polarizations from saturated hysteresis loop oscillograms

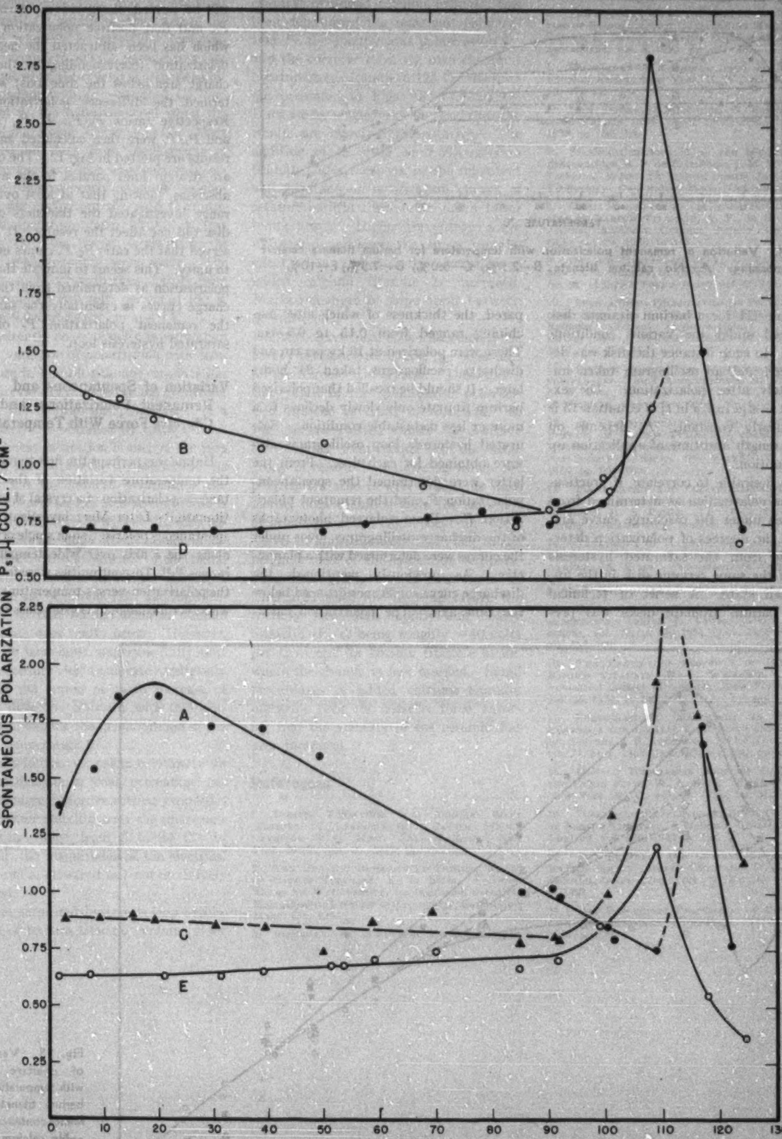


Fig. 13. Variation of spontaneous polarization P_s with temperature for barium titanate ceramics containing: A—No calcium titanate; B—2.5%; C—5.0%; D—7.5%; E—10%

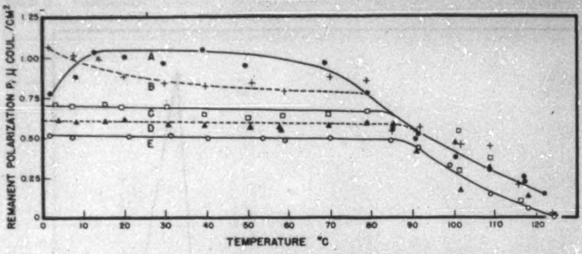


Fig. 14. Variation of remanent polarization with temperature for barium titanate ceramic containing: A—No calcium titanate; B—2.5%; C—5.0%; D—7.5%; E—10%

in Table III for a barium titanate disk polarized under the various conditions given. In each instance the disk was depolarized and an oscillogram taken immediately after polarization. The exponent $-m[d \ln i/d \ln t]$ of equation 13 is very nearly constant. P_s depends on field strength and time of application up to saturation.

It is desirable to correlate, if practicable, the polarization as determined from the area under the discharge curve and one of the degrees of polarization determinable from the saturated hysteresis loop of the same ceramic disk in the unpolarized state. A series of technical grade barium titanate disks was prepared, the thickness of which after machining ranged from 0.15 to 0.5 cm. These were polarized at 16 kv per cm and discharge oscillograms taken 24 hours later. It should be recalled that polarized barium titanate only slowly declines to a more or less metastable condition. Saturated hysteresis loop oscillograms also were obtained for each disk. From the latter were determined the spontaneous polarization P_s and the remanent polarization P_r. From enlarged photographs of the discharge oscillograms, areas under the curves were determined with a planimeter. As previously mentioned, the discharge curve sometimes dropped below the time axis. The polarization calculated from the area above the time axis will be termed, the positive polarization P_0, and that positive polarization from which has been subtracted the negative polarization (corresponding to the discharge area below the time axis) will be termed the difference polarization P_t. Respective ratios P_0/P_s, P_t/P_s, P_0/P_r and P_t/P_r were then calculated and the results are plotted in Fig. 12. The curves are straight lines parallel to the axis of abscissas, showing that at least over the range investigated the thickness of the disk did not affect the results. It is observed that the ratio P_0/P_r comes nearest to unity. This seems to indicate that the polarization as determined from the discharge curves is essentially the same as the remanent polarization P_r of the saturated hysteresis loop.

Variation of Spontaneous and Remanent Polarizations, and Coercive Force With Temperature

Hulme was perhaps the first to measure the temperature variation of the spontaneous polarization of a crystal of barium titanate.[19] Later Merz investigated the spontaneous polarization in single crystals along the c axis over wide temperature ranges.[20,21] Discontinuities were found in the polarization-versus-temperature curve wherever a change in crystal class occurs.

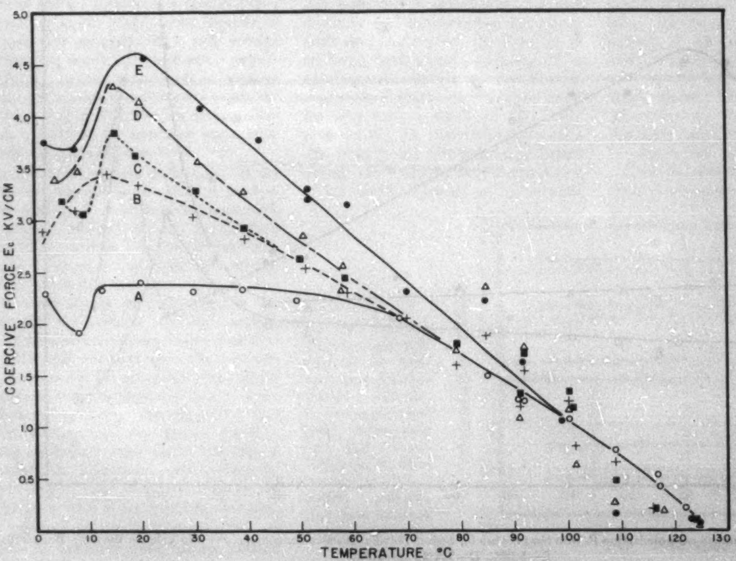


Fig. 15. Variation of coercive force with temperature for barium titanate ceramics containing: A—No calcium titanate, B—2.5%, C—5.0%, D—7.5%, E—10%

This observation is in accordance with the assumption that the spontaneous polarization is parallel to the c axis of the crystal when it is tetragonal (approximately from 0 to 120 C), parallel to a face diagonal when the crystal is orthorhombic (approximately from −80 to 0 C), and parallel to the body diagonal when the crystal is trigonal (below −80 C). Merz deduced his data from oscillograms of hysteresis loops. The spontaneous polarization remains nearly constant (8μ coulombs per cm^2) from the −80 C transition temperature down to 4.2 degrees Kelvin (K), the lowest temperature at which measurements were made. The coercive field strength rises sharply from approximately 250 volts per cm at 180 K to over 9,000 volts per cm at 4.2 K.

Apparently no systematic investigation of the variation of polarization with temperature in barium titanate ceramics has been previously reported in the literature. When the spontaneous polarization of a ferroelectric is obtained by extrapolation of the linear saturation branch of the loop to zero field strength, the assumption is made that the induced polarization depends linearly on the field strength and is given by $P_s=(\epsilon-1)\ E/4\pi$. It also assumes that the slope of this branch is determined by the magnitude of the initial dielectric constant which is measured with the use of a weak field and high frequency so that reorientation of domain structure does not occur. However, Rzhanov has shown experimentally that this procedure yields accurate results only outside the region of rapid increase of initial dielectric constant with temperature, i.e., outside the neighborhood of the Curie temperature.[10]

The addition of calcium titanate to barium titanate in small percentages results in more uniform electrical properties and greater stability over the important operating range from 0 to 40 C. In general, the magnitudes of the electrical constants are lowered but not excessively lowered.

From saturated hysteresis loop oscillograms of barium titanate ceramic disks, containing 0, 2.5, 5.0, 7.5 and 10.0 percentages of added calcium titanate, has been determined the remanent polarization P_r, the spontaneous polarization P_s, and the coercive force E_c, over the temperature range from 0 to 125 C. Results are presented in Figs. 13, 14, and 15. Data are for single disks only and thus the results are merely representative. The addition of as little as 2.5% calcium titanate shifts the peak in the remanent and spontaneous polarization curves of ceramic barium titanate to much lower temperatures. Higher percentages serve to linearize and lower the curves, but the effect is less marked as the percentage of added calcium titanate is increased. Marked changes in slope occur between 80 and 100 C.

The coercive force E_c is a measure of the energy necessary to reverse the polar axis of the ferroelectric crystal. The dipoles that result in the spontaneous polarization become reversed in direction. In Fig. 15 are plotted values for coercive force versus temperature for the same group of ceramics as those on which the polarization studies were made, the data having been obtained from the same lot of oscillograms. Peaks occur in these curves at approximately 15 C, i.e., in the neighborhood of the transition temperature from tetragonal to orthorhombic. Thereafter with rise in temperature the coercive force falls rapidly and nearly linearly, dE_c/dt being roughly −40 volts per C except for barium titanate alone, where the change is less marked. Small percentages of added calcium titanate markedly raise the coercive force, showing that the stability of the ceramic has been increased.

References

1. Domain Formation and Domain Wall Motions in Ferroelectric $BaTiO_3$ Single Crystals, **W. J. Merz.** *Physical Review*, New York, N. Y., vol. 95, 1954, pp. 690–98.

2. The Dynamic Behavior of Domain Walls in Barium Titanate, **E. A. Little.** *Technical Report No. 87*, Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Mass., Oct. 1954.

3. Rochelle Salt as a Dielectric, **C. B. Sawyer, C. H. Tower.** *Physical Review*, New York, N. Y., vol. 35, 1930, pp. 269–73.

4. The Dependence of the Dielectric Properties of Rochelle Salt Crystals on Mechanical Stress, **R. David.** *Helvetica Physica Acta*, Basel, Switzerland, vol. 8, 1935, pp. 431–84.

5. Properties of Rochelle Salt, **H. Mueller.** *Physical Review*, New York, N. Y., vol. 47, 1935, pp. 175–91.

6. New Seignette-Electrics, **G. Busch.** *Helvetica Physica Acta*, Basel, Switzerland, vol. 11, 1938, pp. 269–98.

7. Titania Ceramics II, **A. von Hippel, R. G. Breckenridge, A. P. de Bretteville, J. M. Brownlow.** *Technical Report 11*, Massachusetts Institute of Technology, Cambridge, Mass., Oct. 1945.

8. Oscillograph Study of Dielectric Properties of Barium Titanate, **A. P. de Bretteville.** *Journal*, American Ceramic Society, Columbus, Ohio, vol. 29, 1946, p. 303.

9. On the Hysteresis Loops of $BaTiO_3$ Ceramics, **S. Nomura, S. Sawada.** *Journal*, Physical Society of Japan, Tokyo, Japan, vol. 4, 1949, p. 120.

10. Spontaneous Polarization of Polycrystalline Specimens of Barium Titanate, **A. V. Rzhanov.** *Zhurnal eksperimentalnoi i teoreticheskoi fisiki*, Moscow, USSR, vol. 19, no. 4, 1949, pp. 336–45.

11. Interim Report on Ferroelectric Materials and Their Applications, **H. Diamond, L. W. Orr.** *Technical Report No. 31*, University of Michigan, Ann Arbor, Mich., 1954.

12. An Experimental Study of Polarization Effects in Barium Titanate Ceramics, **T. F. Hueter, D. P. Neukaus, J. Kolb.** *Journal*, Acoustical Society of America, New York, N. Y., vol. 26, 1954, pp. 696–703.

13. Polarization and Electromechanical Behavior of Barium Titanate Ceramics, **S. Levine, H. Blood, D. S. Potter.** *TE 55-17*, 1955, University of Washington, Seattle, Wash.

14. Treatment of Harmonics in Alternating Current Theory by Means of a Harmonic Algebra, **A. Press.** *Publications in Engineering*, University of California, Los Angeles, Calif., vol. 2, no. 5, 1919, p. 139.

15. Harmonic Analysis for Nonlinear Characteristics, **Laurel J. Lewis.** *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 693–700.

16. The Rectifying Property of Polarized Barium Titanate, **D. L. Waidelich.** *Journal*, Acoustical Society of America, New York, N. Y., vol. 25, 1953, p. 796.

17. Ferroelectric Materials as Storage Elements for Digital Computers and Switching Systems, **J. R. Anderson.** *AIEE Transactions*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 395–401.

18. Double Hysteresis Loop of $BaTiO_3$ at the Curie Point, **W. J. Merz.** *Physical Review*, New York, N. Y., vol. 91, 1953, pp. 513–17.

19. Dielectric Properties of Single Crystals of Barium Titanate, **J. Hulme.** *Nature*, London, England, vol. 160, 1947, p. 127.

20. The Electric and Optical Behavior of $BaTiO_3$ Single-Domain Crystals, **W. J. Merz.** *Physical Review*, New York, N. Y., vol. 76, 1949, p. 1221.

21. The Dielectric Properties of $BaTiO_3$ at Low Temperatures, **W. J. Merz.** *Ibid.*, vol. 81, 1951, p. 1064.

Graphical Determination of the Fourier Series Coefficients

W. H. MIDDENDORF
MEMBER AIEE

ALMOST all of the many good textbooks on analysis of electric circuits include analysis of nonsinusoidal waves. Equations are usually given which determine the amplitude of each term of the Fourier trigonometric series. If the dependent variable can be expressed in terms of the independent variable, the familiar relations

$$a_0=\frac{1}{2\pi}\int_0^{2\pi} y\,dx \tag{1}$$

$$a_n=\frac{1}{\pi}\int_0^{2\pi} y\cos nx\,dx \tag{2}$$

$$\text{and } b_n=\frac{1}{\pi}\int_0^{2\pi} y\sin nx\,dx \tag{3}$$

may be used to determine the coefficients.[1] In many situations, however, the dependent variable cannot be expressed in terms of the independent variable except by the trigonometric series which the analysis seeks to establish. In that event the relations just given may be changed to permit graphical integration by the method of rectangular approximation. The coefficients are then given by the equations

$$A_0=\frac{1}{m}\sum_1^m y_k \tag{4}$$

$$A_n=\frac{2}{m}\sum_1^m\left[y_k\cos n\left(k-\frac{1}{2}\right)\frac{2\pi}{m}\right] \tag{5}$$

$$B_n=\frac{2}{m}\sum_1^m\left[y_k\sin n\left(k-\frac{1}{2}\right)\frac{2\pi}{m}\right] \tag{6}$$

where

A_0 = constant term of the series
A_n = coefficient of the nth harmonic cosine term
B_n = coefficient of the nth harmonic sine term
y_k = mid-point value of the function in the kth segment
m = total number of segments

Notice that the coefficients are represented by capital letters in equations 4, 5, and 6, to distinguish them from the coefficients obtained with absolute accuracy using the calculus of infinitesimal differences.

This matter of accuracy in the values of the coefficients as determined by equations 4, 5, and 6 should receive some consideration. The summation of the product $y_k \sin n\,(k-{}^1/_2)\,(2\pi/m)$ requires that the values of the function between the limits of the segment do not differ greatly from the value of y_k. Thus, if the wave is irregular, the number of segments m should be increased if fair accuracy is to be obtained. Under such circumstances it is natural for the investigator to select a large number of segments. Even if the curve is a smooth variation and y_k (the mid-point value) is closely equal to all the values of the function over a large segment, the value of m must be selected with care. The second factor of the product shown, $\sin n\,(k-{}^1/_2)\,(2\pi/m)$, must be a close approximation of all the values of the harmonic sine wave that occur in the segment. Thus, if the same accuracy is to be obtained for the coefficients of the higher harmonics, the value of m must be increased over that used for the fundamental and the lower harmonics. It is true that the importance of accuracy usually decreases as the harmonic number increases, since the amplitudes of the waves usually decrease with an increase in n. Nonetheless, a situation can be created wherein each additional sine and cosine wave added to the series makes the resulting composite wave reproduce the original wave less faithfully; i.e., errors of over 100% (per cent) can occur.

The saw-tooth wave shown in Fig. 1 was analyzed by the exact method using equations 1, 2, and 3, and the coefficients compared to those obtained when the approximate equations 4, 5, and 6 are used. Table I shows that the error in B_{10} when 50 segments are used is about equal to the error in B_5 when 25 segments are used, or the error in B_1 when 5 segments are used. Thus Table I indicates that the number of segments must be increased in proportion to the harmonic number, if all coefficients are to be determined with the same order of accuracy. As a general rule, however, this statement is a bit pessimistic. For the usual irregular type of nonsinusoidal wave it is not necessary to increase the number of segments in direct proportion to the harmonic number in order to maintain the same order of accuracy. An increase in the number of segments results in the mid-point value of each segment being more nearly equal to all the other values of the function within the segment. Thus, as the number of segments is increased, the accuracy of the coefficient improves because of the decrease in the deviation of the function from the value of y_k and the decrease in the deviation of the harmonic sine wave from the value of $\sin n\,(k-{}^1/_2)\,(2\pi/m)$. In the case of the saw-tooth wave the mid-point value is the true average for a segment of any width. Thus, the increase in accuracy with an increase in the number of segments shown in Table I is due mainly to an improvement in the value of $\sin n\,(k-{}^1/_2)\,(2\pi/m)$.

Notice also that if only eight segments are used to determine B_5, Table I shows that the error is 117%. Perhaps few investigators would use only eight segments when analyzing a nonsinusoidal wave, but the information given in the preceding paragraph indicates that a similar error would result even using 32 segments, if the coefficients of the 20th harmonic are to be found.

Another method of graphically determining the Fourier series coefficients, the Fischer-Hinnen method,[2] is usually much less laborious than integration by rectangular approximation. This method requires that a point in the harmonic range be determined beyond which all harmonics may be neglected, which may not be easy. Some indication of this point is made possible by determining the d-c component for different higher-order sets of ordinate sums. When the values so obtained are equal, the higher harmonics probably may be neglected.

The need for a method which accurately and quickly determines the coefficients of the higher harmonic terms is made evident by the preceding discussion of the two methods given in most textbooks. A new method is suggested by a study of equations 2 and 3. It is evident from these relations that the coefficient of each term of the Fourier series is equal to twice the average ordinate of the $y \sin nx$ curve; i.e., a curve formed by multiplying the values of the nonsinusoidal function by the values of a unit sine wave having a frequency equal to that of the harmonic.

Paper 56-684, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted September 21, 1955; made available for printing May 11, 1956.

W. H. Middendorf is with the University of Cincinnati, Cincinnati, Ohio.

Vol.
Pt. 1
1956

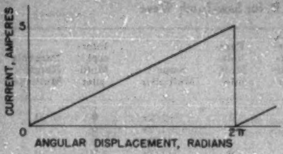


Fig. 1. Saw-tooth wave

Table I. Errors in Fourier Coefficients of Saw-Tooth Wave With Method of Rectangular Summation

Number of Segments	Per-Cent Error in B_1	Per-Cent Error in B_5	Per-Cent Error in B_{10}
50			6.9
40			10.7
25		7.5	
20		11.3	57.0
10		57.3	
8		117	
5	6.8		

Any way in which this average ordinate can be found accurately will permit a correct evaluation of the coefficient. One way to find the average ordinate for a period of the nonsinusoidal wave is to determine the average ordinate of each segment and then find the average of these values. This procedure is made practical by selecting a number of segments which is equal to a multiple of the harmonic number of the coefficient sought. If this is done, literal equations can be developed for the average ordinate of each segment. These equations involve numbers determined by the nonsinusoidal function and multipliers whose values are determined by the particular part of the harmonic wave that exists within the segment under study. These multipliers may be developed in the following manner.

If the equation of the straight line segment shown in Fig. 2 is written as

$$y_1=\alpha x_s+\beta$$

where x_s is measured from the left edge of the segment, and the equation of the harmonic sine segment is written as

$$y_s=\sin(x_s+\theta)$$

then the average ordinate of the product of the straight line and the harmonic sine wave for the segment shown bold in Fig. 2 is

$$0_{1s}=\frac{1}{2\pi/m}\int_0^{2\pi/m}(\alpha x_s+\beta)\sin(x_s+\theta)dx_s$$
$$=\frac{m}{2\pi}\left[\alpha\int_0^{2\pi/m}x_s\sin(x_s+\theta)dx_s+\beta\int_0^{2\pi/m}\sin(x_s+\theta)dx_s\right]$$

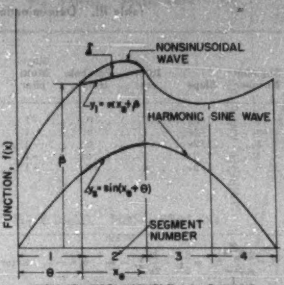


Fig. 2. Typical segments of nonsinusoidal wave and harmonic sine wave

Integrating the first term of the right member by parts yields

$$-\alpha x_s\cos(x_s+\theta)\Big]_0^{2\pi/m}+\alpha\sin(x_s+\theta)\Big]_0^{2\pi/m}$$

The second term of the right member becomes

$$-\beta\cos(x_s+\theta)\Big]_0^{2\pi/m}$$

Thus

$$0_{1s}=\frac{m\alpha}{2\pi}\Big[-x_s\cos(x_s+\theta)+\sin(x_s+\theta)\Big]_0^{2\pi/m}-\frac{m\beta}{2\pi}\Big[\cos(x_s+\theta)\Big]_0^{2\pi/m}=\rho\alpha+\sigma\beta.$$

where

α = slope of the straight line that joins points on the nonsinusoidal wave at the limits of the segment and is related to the scale of the harmonic sought

β = intercept of the straight line at the left hand edge of the segment

ρ = slope-multiplier which depends upon the portion of the harmonic sine term within the segment

σ = intercept-multiplier which is equal to the average of the portion of the harmonic sine wave within the segment

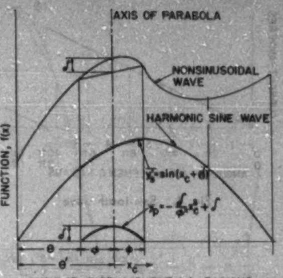


Fig. 3. Parabolic area and harmonic sine wave

The area between the straight line and the function to be analyzed in Fig. 2 is closely represented by the area under a parabola,[3] if the nonsinusoidal function does not have a point of inflection within the limits of the segment. The difference between the area under the nonsinusoidal wave and the trapezoidal area is shown in Fig. 3 for a typical segment. The equation of this parabola may be written as

$$y_p=-\frac{\delta}{\phi^2}x_c^2+\delta$$

if the y-axis is made to coincide with the axis of the parabola. The equation of the sine segment is now written as

$$y_s'=\sin(x_c+\theta')$$

Then the average ordinate of the product of parabola and the harmonic sine wave for the segment shown is

$$0_{ps}=\frac{\delta}{2\phi}\int_{-\phi}^{\phi}\Big[\sin(x_c+\theta')-$$

Table II. Multipliers Which Determine Coefficients of Sine and Cosine Terms

Number of Segments Per Cycle of Harmonic	Segment Number	Limits of Harmonic Within Segment, Degrees	Slope Multiplier	Intercept Multiplier	Parabola Multiplier
Sine Terms					
1	1	0–360	$\rho_{11}=-1$	$\sigma_{11}=0$	$\tau_{11}=0$
2	1	0–180	$\rho_{12}=1$	$\sigma_{12}=0.636$	$\tau_{12}=0.515$
	2	180–360	$\rho_{22}=-1$	$\sigma_{22}=-0.636$	$\tau_{22}=-0.515$
4	1	0–90	$\rho_{14}=1.636$	$\sigma_{14}=0.636$	$\tau_{14}=0.445$
	2	90–180	$\rho_{24}=0.364$	$\sigma_{24}=0.636$	$\tau_{24}=0.445$
	3	180–270	$\rho_{34}=-1.636$	$\sigma_{34}=-0.636$	$\tau_{34}=-0.445$
	4	270–360	$\rho_{44}=-0.364$	$\sigma_{44}=-0.636$	$\tau_{44}=-0.445$
Cosine Terms					
1	1	0–360	$k_{11}=0$	$\lambda_{11}=0$	$\mu_{11}=-0.202$
2	1	0–180	$k_{12}=-0.636$	$\lambda_{12}=0$	$\mu_{12}=0$
	2	180–360	$k_{22}=0.636$	$\lambda_{22}=0$	$\mu_{22}=0$
4	1	0–90	$k_{14}=0.364$	$\lambda_{14}=0.636$	$\mu_{14}=0.445$
	2	90–180	$k_{24}=-1.636$	$\lambda_{24}=-0.636$	$\mu_{24}=-0.445$
	3	180–270	$k_{34}=-0.364$	$\lambda_{34}=-0.636$	$\mu_{34}=-0.445$
	4	270–360	$k_{44}=1.636$	$\lambda_{44}=0.636$	$\mu_{44}=0.445$

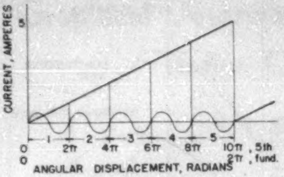


Fig. 4. Saw-tooth wave

$$\frac{1}{\phi^2}x_c^2 \sin (x_c+\theta')\Bigg]dx_c$$

$$=\frac{-\delta}{2\phi}\Bigg[\cos (x_c+\theta')+\frac{2x_c}{\phi^2}\sin (x_c+\theta')-\frac{(x_c^2-2)}{\phi^2}\cos (x_c+\theta')\Bigg]_{-\phi}^{\phi}$$

$$=\tau\delta$$

where δ is defined by Figs. 2 and 3, and τ is the multiplier which depends upon the portion of the harmonic sine wave within the segment.

Likewise, the average ordinate of the product of the straight line and the harmonic cosine wave is

$$0_{1c}=\frac{1}{2\pi/m}\int_0^{2\pi/m}(\alpha x_e+\beta)\cos (x_e+\theta)dx_e$$

$$=\frac{m\alpha}{2\pi}\Bigg[x_e \sin (x_e+\theta)+\cos (x_e+\theta)\Bigg]_0^{2\pi/m}+\frac{m\beta}{2\pi}\Bigg[\sin (x_e+\theta)\Bigg]_0^{2\pi/m}$$

$$=k\alpha+\lambda\beta$$

where

k = slope-multiplier which depends upon the portion of the harmonic cosine term within the segment.

λ = intercept-multiplier which is equal to the average of the portion of the harmonic cosine wave within the segment.

The average ordinate of the product of the parabola and the harmonic cosine wave is

$$0_{pc}=\frac{-\delta}{2\phi}\Bigg[\frac{2x_c}{\phi^2}\cos (x_c+\theta')+\frac{(x_c^2-2)}{\phi^2}\sin (x_c+\theta')-\sin (x_c+\theta')\Bigg]_{-\phi}^{\phi}$$

$$=\mu\delta$$

where μ is the multiplier which depends upon the portion of the harmonic cosine wave within the segment.

The average ordinate of each segment can now be written for the sine and cosine terms as

$$0_s=\rho\alpha+\sigma\beta+\tau\delta$$

and

$$0_c=k\alpha+\lambda\beta+\mu\delta$$

The Fourier coefficients are given in terms of these average ordinates of the segments by the equations

$$A_n=\frac{2}{m}\sum_1^m 0_c \tag{7}$$

and

$$B_n=\frac{2}{m}\sum_1^m 0_s \tag{8}$$

The values of the multipliers have been computed for 1, 2, and 4 segments per cycle of the harmonic sine and cosine waves. Table II presents these values. The method proposed by this paper becomes more attractive when one notices that many of the multipliers are zero when using one or two segments per cycle of the harmonic sought. The multipliers are given two subscripts. The first indicates the segment for which the multiplier is appropriate. The second indicates the total number of segments per cycle of the harmonic.

The method of analysis proposed herein is most easily taught by three examples. The saw-tooth wave previously analyzed by the rectangular summation method will serve as the first example. The coefficient of the fifth harmonic sine term will be determined. Fig. 4 shows the saw-tooth wave of Fig. 1, but the harmonic wave whose coefficient is sought is drawn on the same time axis for ready reference. Of course, this wave cannot be drawn with the correct amplitude. Table III shows the calculations necessary to determine B_5. Notice that the slope is reckoned on the basis of the angular displacement of the harmonic sought. For irregular waves the slope is easily determined by use of a bevel protractor. The tangent of the angle, so measured, times the ratio of the ordinate scale factor, divided by the abscissa scale factor, is equal to the slope.

For the second example, the square wave shown in Fig. 5 will be analyzed to determine the coefficient of the fifth harmonic cosine term. Table IV shows the calculations necessary to determine A_5. Notice that it is possible to use relatively large segments (one per cycle of harmonic) over part of the nonsinusoidal wave and smaller segments where the

Table III. Determination of B_5 for Saw-Tooth Wave

Segment Number	Slope	Intercept	Parabola Height	Slope-Multiplier	Intercept-Multiplier	Parabola-Multiplier	Slope × Multiplier	Intercept × Multiplier	Parabola Height × Multiplier
1	$1/2\pi$	0	0	−1	0	0	$-1/2\pi$	0	0
2	$1/2\pi$	1	0	−1	0	0	$-1/2\pi$	0	0
3	$1/2\pi$	2	0	−1	0	0	$-1/2\pi$	0	0
4	$1/2\pi$	3	0	−1	0	0	$-1/2\pi$	0	0
5	$1/2\pi$	4	0	−1	0	0	$-1/2\pi$	0	0

$$\sum_1^5 0_s=-\frac{5}{2\pi}+0 \quad + \quad 0$$

$$B_5=\frac{2}{m}\sum_1^m 0_s, =\frac{2}{5}\left(-\frac{5}{2\pi}\right), =-\frac{1}{\pi};\ b_5=-\frac{1}{\pi}. \quad \text{Error}=0\%.$$

Table IV. Determination of A_5 for Square Wave

Segment Number	Slope	Intercept	Parabola Height	Slope-Multiplier	Intercept-Multiplier	Parabola-Multiplier	Slope × Multiplier	Intercept × Multiplier	Parabola Height × Multiplier
1	0	5	0	0	0	−0.202	0	0	0
2a	0	5	0	0.364	0.636	0.445	0	$\frac{2\times5\times0.636}{4}$	0
2b	0	−5	0	−1.636	−0.636	−0.445	0		0
2c	0	−5	0	−0.364	−0.636	−0.445	0		0
2d	0	−5	0	1.636	0.636	0.445	0		0
3	0	−5	0	0	0	−0.202	0	0	0
4a	0	−5	0	0.364	0.636	0.445	0	$\frac{2\times5\times0.636}{4}$	0
4b	0	−5	0	−1.636	−0.636	−0.445	0		0
4c	0	−5	0	−0.364	−0.636	−0.445	0		0
4d	0	5	0	1.636	0.636	0.445	0		0
5	0	5	0	0	0	−0.202	0	0	0

$$\sum_1^5 0_c=0 \ + \ 5\times0.636 \ + \ 0$$

$$A_5=\frac{2}{m}\sum_1^m 0_c, =\frac{2}{5}(5\times0.636), =1.272;\ a_5=1.272. \quad \text{Error}=0\%.$$

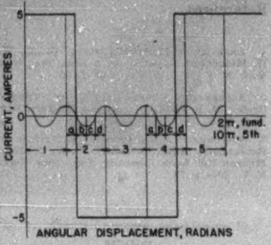


Fig. 5 (left). Square wave

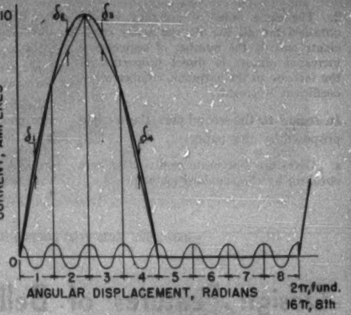


Fig. 6 (right). Half-rectified sine wave

irregularity of the wave demands. When a segment is broken into smaller parts, the average of the ordinates of the parts is equal to the average ordinate of the segment.

The last example will be the analysis of the half-rectified sine wave shown in Fig. 6. The coefficient of the eighth harmonic cosine term will be determined. The necessary calculations are shown in Table V. Notice that for nonsinusoidal waves without discontinuities the sum of the slopes will equal zero. Thus, if one segment per cycle of the harmonic sought is used, the sum of the slopes times the multipliers will equal zero for both sine and cosine coefficients.

These three examples show the basic calculations necessary to analyze a wave by this method. Just as with these examples, most nonsinusoidal waves will not require that all three types of calculations be made. In that event it is not necessary to include the columns of zeros as was done here. Since the occurrence of zeros is the rule rather than the exception, the investigator will do well to determine how many calculations can be avoided before the analysis begins.

Besides its use in determining the coefficients this method permits generalizations to be made concerning the harmonic content of nonsinusoidal waves. This can be of advantage in determining the point in the harmonic range beyond which all harmonics may be neglected. Most important of these generalizations are:

1. A wave that can be faithfully represented by one segment per cycle of the harmonic sought will not yield a coefficient for the sine term of that harmonic unless the wave contains at least one finite jump.

2. A wave that can be faithfully represented by one segment per cycle of the harmonic sought will yield a coefficient for the cosine term of that harmonic proportional to the algebraic sum of the parabola heights divided by the number of segments.

It is true that the waves analyzed in this paper do not require the use of graphical methods. These nonsinusoidal waves were chosen for graphical analysis because the coefficients so obtained can be compared to those known from exact analysis. The exact value of the coefficient is given in the appropriate table (see the coefficient represented by the lower case letter) as well as the error in the value determined the method of this paper. The maximum error of 3% is well within the error to be expected when values are read from a graph.

Although this method of analysis is valid and accurate for the determination of the coefficients of the lower order harmonics, these situations usually require that so many segments be used for each cycle of the harmonic that the computations become excessive. For any nonsinusoidal wave the number of segments chosen must be large enough that the function does not have points of inflection within the segments. The author has used this method for problems with as many as eight segments per cycle of the harmonic sought (the second). The error of the result was less than 3%. However, similar accuracy was obtained by the method of rectangular approximation with a comparable amount of work using 36 segments. The Fischer-Hinnen method also gave comparable accuracy with much less effort. It therefore seems reasonable to propose that the method presented in this paper be used to determine the coefficients of the higher order harmonics and the Fischer-Hinnen method be used for the coefficients of the lower order harmonics.

Conclusions

This paper has given the results of a study of the Fourier series coefficients. The study was twofold. It has been concerned with:

1. The accuracy of the coefficients when found by the standard method of graphical analysis and,

2. A new method of determining the coefficients.

In regard to the first part, it has been found that:

1. Errors of over 100% can result when using the standard method of graphical analysis.

Table V. Determination of A_8 for Half-Rectified Sine Wave

Segment Number	Slope	Intercept	Parabola Height	Slope-Multiplier	Intercept-Multiplier	Parabola-Multiplier	Slope × Multiplier	Intercept × Multiplier	Parabola Height × Multiplier
1	1.12	0	0.35	0	0	−0.202	0	0	−0.071
2	0.47	7.07	0.68	0	0	−0.202	0	0	−0.137
3	−0.47	10.00	0.68	0	0	−0.202	0	0	−0.137
4	−1.12	7.07	0.35	0	0	−0.202	0	0	−0.071
5	0	0	0	0	0	−0.202	0	0	0
6	0	0	0	0	0	−0.202	0	0	0
7	0	0	0	0	0	−0.202	0	0	0
8	0	0	0	0	0	−0.202	0	0	0
							$\sum_{1}^{8} 0_c = 0$	+ 0	−0.416

$$A_8 = \frac{2}{m}\sum_{1}^{m} 0_c, = \frac{2}{8}(-0.416), = -0.104; \quad a_8 = -0.101. \quad \text{Error} = 3\%.$$

2. The same order of accuracy will be obtained for all the Fourier series coefficients only if the number of segments is increased almost in direct proportion to the increase in the harmonic number whose coefficient is sought.

In regard to the second part, the method proposed by this paper:

a. Gives more accurate results than were obtained by the standard method.

b. Requires fewer calculations for the three nonsinusoidal waves analyzed.

c. Suggests generalizations that permit rapid estimations of the amplitudes of the harmonics.

It must be remembered, however, that this method loses many of its advantages when more than two segments per cycle of the harmonic sought are used to fit the straight lines and parabolas to the irregular variations of the nonsinusoidal wave.

References

1. Analysis of Electric Circuits (book), W. H. Middendorf. John Wiley & Sons, Inc., New York, N. Y., 1956, articles 17.03, 17.05.

2. Alternating Current Circuit Theory (book), M. B. Reed. Harper and Brothers, New York, N. Y., 1948, pp. 237–43.

3. Mathematical and Physical Principles of Engineering Analysis (book), W. C. Johnson. McGraw-Hill Book Company, Inc., New York, N. Y., 1944, p. 78.

Design Features of Bell System Wire Spring Relays

H. M. KNAPP
NONMEMBER AIEE

THE Bell Telephone System wire spring relays have become the basic circuit components of telephone switching systems developed during the past few years. These relays are provided in many varieties to meet circuit needs. They are similar in operation and construction and have been described to a large extent in earlier publications.[1,2] It is the purpose of this paper to highlight some of the design features which permit automatic control, during manufacture, of important operating characteristics. This subject has not been covered in the earlier articles and is of considerable importance to low-cost manufacture and proper operation of these relays.

While control during manufacture of some of the operating characteristics is also true of other relays, the aim in the development of wire spring relays was to attain essentially all desired characteristics automatically without hand adjustment. This provides a more uniform product at low cost but calls for special design and manufacturing features. Before the discussion of these features, a brief description of the three classes or families of wire spring relays is in order.

The General-Purpose Relays

The general-purpose family of relays has now grown to more than 200 codes even though it is the continuing objective to keep the codes to a minimum. However, this number is justified economically because of the variety of switching requirements encountered. These requirements call for contact combinations ranging between 1 and 24 contacts per relay, high-speed operation in some cases, slow operation in other cases, and applications requiring high sensitivity, marginal operation or unusually long life—up to a billion operations.

This variety of operating conditions is accomplished with relays having basic parts which are similar and, in some cases, do not require additional tools and machines. For example, Fig. 1 shows a relay with 12 make and 6 break contacts. The light armature, which is designed for speed, limits the number of contacts which can be operated to a maximum of 18. For added pull, a heavier armature is employed and is shown on a relay equipped with 24 make contacts in Fig. 2. Fig. 3 illustrates a relay with relatively few contacts and a buffer spring. This spring is used to control the release current of relays where special release requirements must be met. These three cases illustrate that, by minor differences in some parts or by using simple additional parts, a wide variety of operating conditions can be met. Hence the term general purpose is applied to this class of relays.

The Multicontact Wire Spring Relays

The multicontact relays are similar in construction in most respects to the general-purpose relays. They are larger because they operate more contacts and must meet equipment requirements as to spacing of terminals and multiple strapping of wires between relays. They perform the special function in telephone switching of connecting a large group of leads from one switching frame to another. In new systems they are used in units of 30 contacts. One of these is shown in Fig. 4. They are faster in operation than the former flat spring multicontact relays and they have much longer life. They are also designed as an electrical and mechanical replacement for the former relays. For these applications, two 30-contact units are mounted on a common frame. One of these relays is shown in Fig. 5.

The Dual Relays

The dual relays are the third family of wire spring relays. They are a recent extension of the wire spring relay principles.

Each dual relay consists of two independent relays mounted on the same structure; see Fig. 6. They are designed to make more efficient use of space and materials where no more than five transfer contact combinations are required for each relay. They employ the same wire blocks, and one dual unit occupies the same mounting space as a single general-purpose relay. The greater economy of the dual relay within its field is made obvious by Fig. 7 which shows the parts needed to do the same job with both types of relays.

Automatic Control of Operating Characteristics—Spring Forces

Most of the important operating characteristics in all of these relays such as speed of operation, contact reliability,

---

Paper 56-662, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 16, 1956; made available for printing May 1, 1956.

H. M. Knapp is with Bell Telephone Laboratories, Inc., New York, N. Y.

The author wishes to acknowledge the co-operation and help of the engineers of the Western Electric Company who are responsible for the manufacture of the Bell System wire spring relays and who contributed in an important way to many of the features described in this paper, and of A. L. Jeanne who assisted in the preparation of the text and the illustrations.

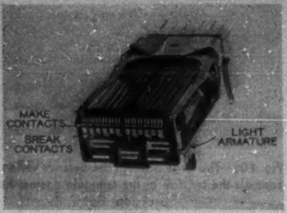


Fig. 1. General-purpose relay using light armature

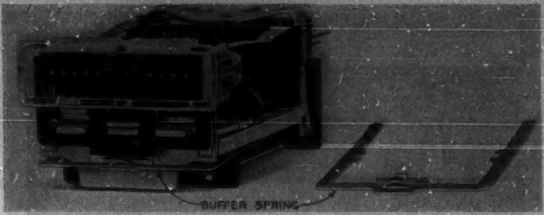


Fig. 3. General-purpose relay using buffer spring

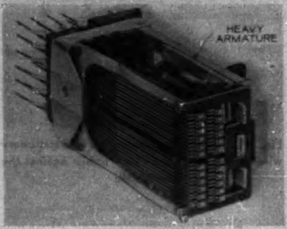
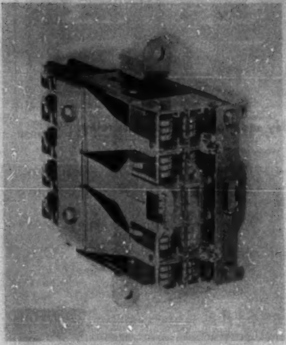


Fig. 2. General-purpose relay using heavy armature

Fig. 4 (right). Muticontact relay for new applications



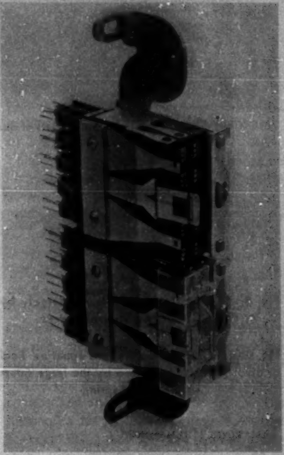


Fig. 5. Multicontact replacement for former multicontact relays; 287-type relay with one contact cover removed

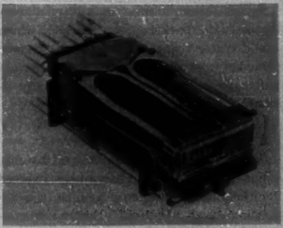


Fig. 6. Dual relay (two in one)

etc., relate directly to forces developed by cantilever springs, to armature travel and to the contact closure points during the armature stroke. Allowances must be made in establishing their magnitude, for variations in manufacture and for changes as a result of aging and use. The discussion which follows will be limited essentially to design features which permit automatic control of forces, travel, and contact closure without resorting to hand adjustment in manufacture.

Of these three, the spring forces are the easiest to obtain automatically. These are attained by machine-bending of the springs prior to assembly in the relay. When assembled, the proper tension is obtained by reason of the deflection caused by this bend. The stiffness of the spring should be low so that the amount of deflection can be as large as safe stresses and other considerations permit in order that manufacturing variations in offset, spring thickness, etc., will have little effect on the desired force. Several examples are cited to illustrate this principle as applied to the design of wire spring general-purpose relays. The other types of wire relays employ similar features.

CONTACT FORCE

The magnitude of the contact force contributes to contact reliability of the relay and to the reduction of contact chatter. The lower the force the greater is the likelihood of failure caused by dust, and of chatter or momentary contact openings caused by spring vibration. While high contact force contributes to better contact performance, more power is needed to operate the relay. Hence good control in manufacture of an optimum value of contact force is important.

Contact force for both make and break contacts is governed almost entirely by the large bend formed in the movable wires during manufacture. The movable wires and the effect of this bend before and after assembly of the wire block into the relay are illustrated in Fig. 8. The force developed by the wire is applied against the actuating card when the contact pair is open and against the stationary contact when the contact pair is closed. Hence, this deflection provides the contact force. With the use of low-stiffness wire, which permits a large deflection, the normal manufacturing variations in the amount of bend, in wire diameter, etc., and variations resulting from wear in service have only a minor effect on the total force developed. Hence, a well-controlled contact force is assured not only in manufacture but also during the service life of the relay.

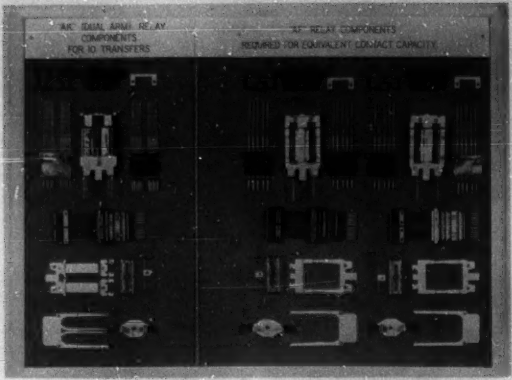


Fig. 7. Comparison of parts of dual relay and two general-purpose relays

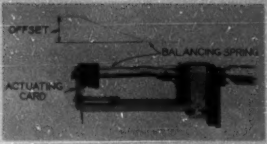


Fig. 10. The large offset in the balance spring controls the tension of the armature against its backstop

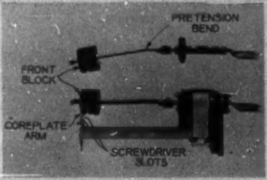


Fig. 11. The pretensioning of the stationary wires holds the front block firmly against the core plate

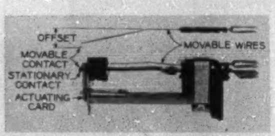
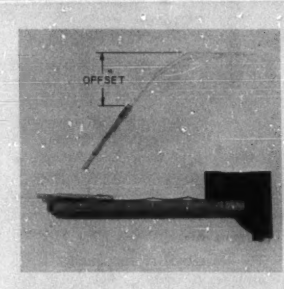


Fig. 8. Contact force is governed largely by the offset in the movable wires

Fig. 9 (right). The force at the armature heel or pivot point is governed by the large offset in the hinge spring



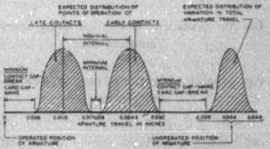


Fig. 12. Components of armature travel on a relay with sequence contacts

Note: No hand adjustment for: contact interval, individual contact closure, and armature travel

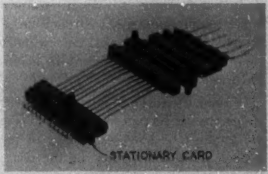


Fig. 13. Stationary wire block assembly

ARMATURE HEEL FORCE

The armature heel force is that of the armature legs against the core at the pivot point. This force should never reach zero during the life of the relay and should be substantially higher to minimize the bounce of the armature against its backstop on release which could cause reoperation of the contacts. The large predeflection of the armature hinge spring governing this force is illustrated in Fig. 9.

BALANCING SPRING FORCE

This is the force, also obtained by machine-bending before assembly, exerted by the balancing spring against the actuating card. It holds the make contacts open and the armature against its backstop with the relay unoperated. It guarantees the release of the relay and supports the card. Its application is illustrated in Fig. 10.

STATIONARY WIRE BLOCK TENSION

The front block of the stationary wires exerts a force against each core plate arm by reason of the machine-made pretension bend in the wires. This is illustrated in Fig. 11. The control of this force is important because its magnitude determines whether the block is held firmly in position against vibration so that wear between the block and the core plate does not occur. Wear at these surfaces could cause objectionable changes in operating characteristics.

Despite the importance of all of these forces in the proper functioning of the relay, their attainment, automatically, by large predeflections made it unnecessary in most cases to specify direct requirements, to hand-adjust, or even to inspect. This results in a more uniform product at lower cost. The only exception applies to the balancing spring which requires some adjustment on relays where the range of variation must be reduced to meet close current flow or armature back tension requirements.

Short Armature Travel

The importance of making armature travel as short as possible, consistent with proper operation of the relay, is obvious to anyone familiar with relay design. Smaller armature movement results in shorter operating and releasing times, greater operating sensitivity, longer life, and less chatter of the contacts.

To avoid the hazards inherent in short travel, the factors governing this must be controlled as closely as possible. The factors to be discussed herein are the distribution of the points of operation of the contacts, the interval of sequence

Vol. 7
Pt. 1
1956

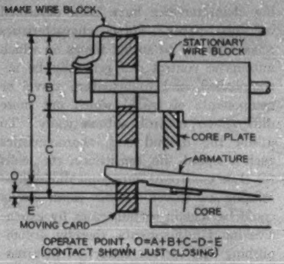

MAKE WIRE BLOCK
STATIONARY WIRE BLOCK
A
B
D
C
O
E
CORE PLATE
ARMATURE
CORE
MOVING CARD
OPERATE POINT, O=A+B+C−D−E
(CONTACT SHOWN JUST CLOSING)

Fig. 14. Important dimensions governing make contact operate point on general-purpose relays

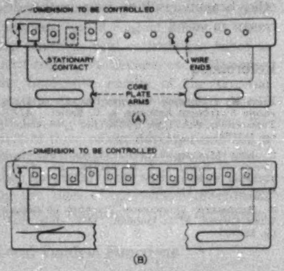

DIMENSION TO BE CONTROLLED
STATIONARY CONTACT
WIRE ENDS
CORE PLATE ARMS
(A)
DIMENSION TO BE CONTROLLED
(B)

Fig. 15(A) Possible effect of variables on contact location. (B). These effects eliminated by functional location of contacts

contacts, and the distribution of total travel itself, as shown in Fig. 12. Close control of these characteristics in manufacture means narrower distributions and, hence, shorter travel is permitted without upsetting design margins for wear, atmospheric changes, etc.

To obtain close control in these cases, the design objective was to keep the number of variables and dimensions governing these characteristics to a minimum. It was also recognized that the design should lend itself to the type of manufacturing operations in which close control of the essential dimensions would be possible. For example, a rolling or a swaging operation provides closer control of a dimension than one governed by a forming operation.

Distribution of Contact Closure Point

The closure point of the contacts in the stroke of the armature is most difficult to control automatically within limits comparable to those obtained usually by hand adjustment. This is because of the number of dimensions that would normally affect this characteristic. Also the large number of contacts which these relays are designed to operate complicates this problem. However, the fact that one contact of each pair is stationary facilitates the use of design and manufacturing techniques to reduce the number of variables. It permits the stationary card, usually referred to as the front molded block, to be an integral part of the stationary wire assembly, as shown in Fig. 13. This led to a new manufacturing technique which positions the contacts in welding where they need to be in the assembled relay.

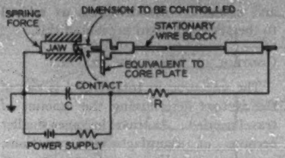

SPRING FORCE
DIMENSION TO BE CONTROLLED
JAW
STATIONARY WIRE BLOCK
EQUIVALENT TO CORE PLATE
CONTACT
C
R
POWER SUPPLY

Fig. 16. Schematic of circuit for percussive welding of stationary contacts

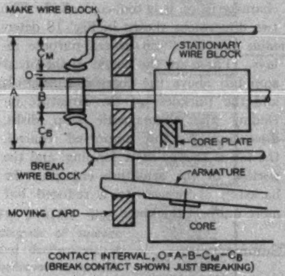

MAKE WIRE BLOCK
STATIONARY WIRE BLOCK
C_M
O
A
B
C_B
CORE PLATE
BREAK WIRE BLOCK
ARMATURE
MOVING CARD
CORE
CONTACT INTERVAL, $O=A-B-C_M-C_B$
(BREAK CONTACT SHOWN JUST BREAKING)

Fig. 17. Important dimensions controlling break–make contact interval

The location of these contact surfaces is an important dimension controlling the contact closure point. The several primary dimensions governing this characteristic, including the location of the stationary contacts, dimension *B*, are illustrated in Fig. 14. If the stationary contacts were attached to the wires by the usual welding methods, their location would be affected by additional factors. The effect of these factors are shown in Fig. 15(A) and are caused by:

1. Variations of wire position in the block.
2. Warpage of the block in molding.
3. Deflection of the block caused by the tension of the wires holding it down.
4. Deflection of the wires caused by the pressure of the welding electrode.

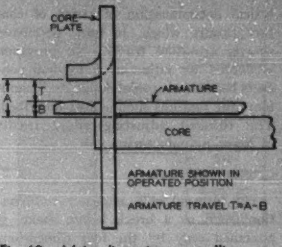

CORE PLATE
A
T
B
ARMATURE
CORE
ARMATURE SHOWN IN OPERATED POSITION
ARMATURE TRAVEL T=A−B

Fig. 18. Major dimensions controlling armature travel on general-purpose relays

These variables are eliminated by locating the contacts in the process of welding, relative to the front mounting surfaces of the stationary wire assembly, with the use of a special welding process. This is done while applying the same forces on the part in the welding fixture as are applied in the assembled relay. Fig. 15(B) illustrates the good alignment of the contact surfaces obtained by this method.

To accomplish this result, a special welding process was developed by the Western Electric Company.[9] This process, which is a special form of percussive or arc welding, permits one welding electrode to be connected to the terminal end of the wire, well away from the weld zone. The other electrode is in the form of a movable jaw which holds the contact being welded in a fixed plane relative to the front mounting surfaces of the stationary wire assembly. This is shown schematically in Fig. 16. With the usual welding, both electrodes must be connected adjacent to the weld zone, which results in deflection of the wire ends and poor control in the location of the contact surfaces.

Even with the elimination of variables as just described, the margin in the automatic control of contact closure did not appear adequate since it might not always be possible to meet all the desired limits on the controlling dimensions. Because of this, a means for making a gang adjustment was provided in the relay. The front-molded block is supported on the two adjustable arms of the core place; see Fig. 11. These arms can be bent with the use of a screwdriver inserted in the slot so as to alter the position of the fixed contacts as a group. By this means, manufacturing variations common to all the contacts, such as the variations of dimensions *C* and *E* of Fig. 14, can be removed. Hence, such variations were excluded in the analytical determination of range of contact closure which, in turn, permitted shorter armature travel. No provision is needed for adjusting individual contacts because of special design and manufacturing features such as just described. The reduced adjusting effort results in lower manufacturing cost.

Contact Interval

The interval of a sequence combination, such as a break before make, is governed only by the four dimensions shown in Fig. 17. Dimension A, the card dimension controlling the separation of the movable wires, is obtained by punching and shaving operations. Dimension B, the height of the stationary contact, is controlled largely by a rolling operation in the manufacture of the contact tape. This contact is sheared from tape consisting of a layer of palladium on top and another layer on the bottom of a nickel base. Dimensions C_m and C_b, the height of the offset and the contacts of the movable wires, depend upon the accuracy of a swaging or coining operation. In this latter case, it would have been better to have used straight instead of offset wires for the sake of simplicity and control. However, space and assembly considerations as related to the actuating card prevented this.

Control of Armature Travel

The variation in total travel is one of the factors determining the amount of travel needed. As travel becomes smaller because of manufacturing variations, contact separation of a make contact, for example, may be seriously reduced.

The travel on wire spring relays is not adjustable. Like most of the other characteristics, it is tool-controlled. The two dimensions shown in Fig. 18 determine the travel on general-purpose relays. These are the height of the backstop tab above the core, dimension A, and the thickness of the armature including the nonmagnetic stop disks, dimension B. For improved control, the tab is shaved after forming, and the portion of the armature engaging the backstop is swaged to a reduced but more uniform thickness.

Summary

Special design features of the Bell Telephone System wire spring relays which contribute to automatic control, during manufacture, of important operating characteristics have been described in this paper. These characteristics relate directly to spring forces on which automatic control can generally be accomplished, and to armature travel or components of travel which are more difficult to control automatically. To attain the desired end, characteristics such as speed, life, sensitivity, etc., calls for close control of important dimensions relating to travel. While closer limits could be met in certain cases by expensive hand adjustments, thereby accomplishing still further reduction in armature travel, it is believed that the best economic balance has been achieved in these relays between manufacturing cost and value of additional speed, life, and other benefits resulting from further decreases in travel.

References

1. A New General-Purpose Relay for Telephone Switching Systems, A. C. Keller. *AIEE Transactions*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 413–28.

2. New Multicontact Relay for Telephone Switching Systems, I. S. Rafuse. *Bell System Technical Journal*, New York, N. Y., vol 33, Sept. 1954, pp. 1111–32.

3. Automatic Percussion Welding of Single Wire Combs, A. L. Quinlan. *Ibid.*, July 1954, pp. 885–95.

The Full Stature of the Crossbar Tandem Switching System

J. MESZAR
MEMBER AIEE

CROSSBAR tandem is a relatively young switching system. It first went into service in 1941. Its original role was well defined although, in retrospect, severely limited; it was to take over in large metropolitan areas the functions of the panel-type office selector tandem system, even as the no. 1 crossbar system[1] took over from the local panel central office. However, even at that early date, its latent potentialities for a broader field of usefulness were appreciated. This appreciation could not then be implemented because during World War II development of telephone switching facilities at Bell Telephone Laboratories came to a near standstill. The needs of national defense transformed switching development engineers almost overnight into designers and teachers of radar, sonar, etc. True, there followed the great postwar surge of switching development activities but these, in turn, involved new systems urgently needed to start customer and operator nation-wide dialing.[1] This period saw the development of the no. 5 crossbar local switching system,[2] the automatic message accounting (AMA) system,[3] and the *4A* crossbar long-distance switching system.[4] In view of this tremendous development program of new systems, the available development talent had to be practically rationed, and so only a moderate amount of effort could be applied to the crossbar tandem system to increase its scope of usefulness. The outcome of this effort and the basic description of the crossbar tandem system were presented in 1950.[5] The closing paragraph of this 1950 paper was the following statement: "An extensive development program is now in progress to incorporate new features which will still further enhance the usefulness of crossbar tandem in the rapid expansion of automatic switching in the Bell Telephone System." By now, this extensive development program referred to is nearing completion, and it is the purpose of the present paper to highlight the significant new features embodied in the 1956 design version of the crossbar tandem system.

It is well to bring out that these new features have not been achieved by any great technological changes in the crossbar tandem system. The 1956 design version of the system remains a typical electromagnetic switching system based on crossbar switches and relays, as was the 1941 version. Here and there, of course, entirely new devices and circuit configurations have been incorporated,

---

Paper 56-665, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted April 2, 1956; made available for printing April 23, 1956.

J. Meszar is with Bell Telephone Laboratories, Inc., New York, N. Y.

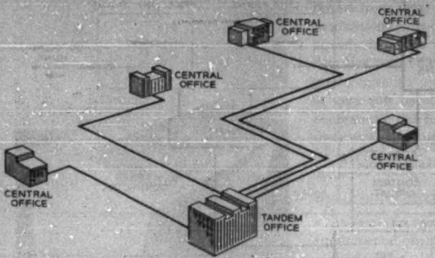

CENTRAL OFFICE
CENTRAL OFFICE
CENTRAL OFFICE
CENTRAL OFFICE
CENTRAL OFFICE
TANDEM OFFICE

Fig. 1 (above). Local tandem switching

in accordance with the over-all progress of electromechanical switching technology. Examples of such innovations are the use of wire spring relays for general control functions, sealed contact relays as memory elements, cold cathode gas tubes as pulse-responsive devices, etc. However, the impressive results of the crossbar tandem developments reported herein are not in this direction. The true impact of these recent developments lies in the great increase in the functional versatility and level of responsibility possessed by the 1956 design version of crossbar tandem as compared to its earlier version.

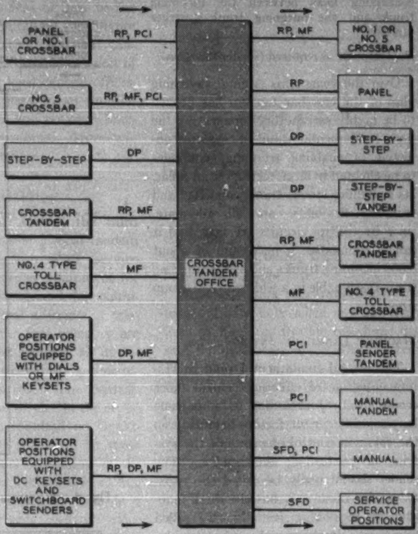

PANEL OR NO. 1 CROSSBAR
RP, PCI
NO. 5 CROSSBAR
RP, MF, PCI
STEP-BY-STEP
DP
CROSSBAR TANDEM
RP, MF
NO. 4 TYPE TOLL CROSSBAR
MF
OPERATOR POSITIONS EQUIPPED WITH DIALS OR MF KEYSETS
DP, MF
OPERATOR POSITIONS EQUIPPED WITH DC KEYSETS AND SWITCHBOARD SENDERS
RP, DP, MF
CROSSBAR TANDEM OFFICE
RP, MF
NO. 1 OR NO. 5 CROSSBAR
RP
PANEL
DP
STEP-BY-STEP
DP
STEP-BY-STEP TANDEM
RP, MF
CROSSBAR TANDEM
MF
NO. 4 TYPE TOLL CROSSBAR
PCI
PANEL SENDER TANDEM
PCI
MANUAL TANDEM
SFD, PCI
MANUAL
SFD
SERVICE OPERATOR POSITIONS

Fig. 3 (right). Block diagram of sources and destinations of traffic routed through crossbar tandem systems

Local Tandem Functions

As was mentioned, the original mission of crossbar tandem was to take over the functions of the local panel office selector tandem system. In accordance with the pattern of Fig. 1, the basic need for a local tandem switching system arises in large metropolitan areas where local telephone service is furnished by dozens or even hundreds of central offices. In such areas tandem systems are used:

1. To effect trunking economies by combining small amounts of interoffice traffic originating in the various central offices and routing this combined traffic over a common trunk group to the required destination, thus eliminating the necessity for inefficient, small, direct interoffice trunk groups.

2. To effect central office economies by serving as a machine language interpreter between the different types of central offices, thereby eliminating the costly necessity of equipping every central office for direct communication with every other type of central office.

3. To effect over-all service economies by centralizing equipment for rendering 1-way message service (weather reports, etc.) and equipment needed as normal adjuncts of telephone service (charging equipment, etc.). Centralization results in efficient use of such equipment as compared with furnishing the same at each local central office requiring it.

TANDEM TRUNK SWITCHING

The basic structure of the crossbar tandem system was devised to fulfill eminently these fundamental objectives of a local tandem switching system. Fig. 2 is an over-all functional block diagram of the crossbar tandem system.

As indicated on this figure, the switching network of the system consists of the by now well-known crossbar link frames. Incoming trunks from the calling central offices appear on trunk link frames and outgoing trunks to the called central offices appear on office link frames. Through a system of links and junctors, which interconnect the crossbar switches on these frames, any incoming trunk can be connected to any outgoing trunk. For each connection there are several possible paths (or channels) through these frames and, on a call, one of these paths is selected by the common control equipment.

To establish a path, the incoming trunk calls for the services of the common control equipment. This equipment consists of a small number of elements (senders, markers, etc.), each of which is called into service for a relatively short time, performs its functions, releases, and is free to serve another call. The basic functions of these common control elements are:

1. Store the digits as they are received (sender function).

2. Translate these digits into switching control information (marker function).

3. Test for and select an idle outgoing trunk (marker function).

4. Test for and select a matching set of

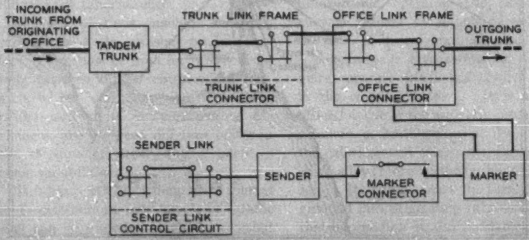

INCOMING TRUNK FROM ORIGINATING OFFICE
TANDEM TRUNK
TRUNK LINK FRAME
TRUNK LINK CONNECTOR
OFFICE LINK FRAME
OFFICE LINK CONNECTOR
OUTGOING TRUNK
SENDER LINK
SENDER LINK CONTROL CIRCUIT
SENDER
MARKER CONNECTOR
MARKER

Fig. 2. Functional block diagram of crossbar tandem

connecting links between the incoming trunk and the outgoing trunk (marker function).

5. Outpulse as required (sender function).

Over the years, this original switching plan of the crossbar tandem has proved to be highly satisfactory for meeting the first fundamental tandem system objective of enabling trunking economies to be effected in large metropolitan areas. Its switching speed, traffic capacity, and trunking flexibility are all adequate, since its control circuits are fast, and it can accommodate a maximum of about 3,000 incoming trunks and 4,000 outgoing trunks assignable in a wide range of group sizes.

Machine Language Interpreting

The second fundamental tandem system objective of effecting central office economies by functioning as a machine language interpreter focuses attention on the common control elements of the system, in particular, the sender. The sender should really be called the register-sender because it has the dual function of registering initially the called telephone number and, subsequently, after the marker has established the connection through the office, of sending information forward to the succeeding office.

The senders are the natural units of the tandem system for acting as machine language interpreters. Senders can be readily designed to receive digits in the language of the calling central office, and to send forward digits in a different language as required by the called central office.

One of the measures, then, of the full stature acquired by the cross-bar tandem system is the versatility of its senders. The preceding system, i.e., the panel sender tandem system, could only receive digits in one machine language called the panel call indicator (PCI). In this language, the tandem office receives from the originating central office a set of timed code pulses comprising different magnitudes and polarities of direct current. The 1956 design version of crossbar tandem has this same ability of receiving PCI pulses, but it also has a greatly increased competence along these lines. It has the ability to receive digits from calling central offices in three additional machine languages:

1. In the dial pulsing (DP) language, in which digits come to the tandem in terms of simple decimal ON-OFF pulses of direct current.

2. In the multifrequency (MF) language, in which the digits come to the tandem in

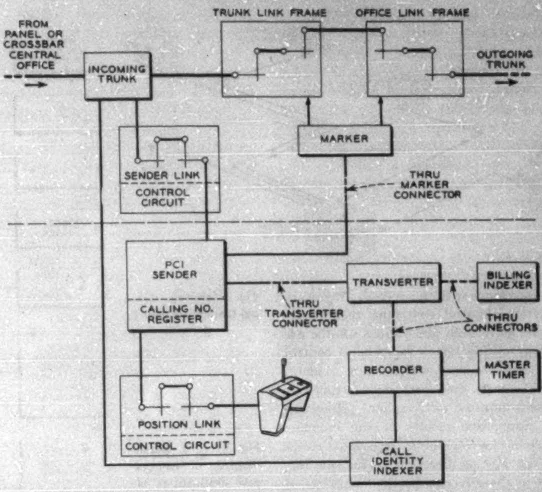

FROM PANEL OR CROSSBAR CENTRAL OFFICE
TRUNK LINK FRAME
OFFICE LINK FRAME
INCOMING TRUNK
OUTGOING TRUNK
MARKER
SENDER LINK
CONTROL CIRCUIT
THRU MARKER CONNECTOR
PCI SENDER
CALLING NO. REGISTER
TRANSVERTER
BILLING INDEXER
THRU TRANSVERTER CONNECTOR
THRU CONNECTORS
RECORDER
MASTER TIMER
POSITION LINK
CONTROL CIRCUIT
CALL IDENTITY INDEXER

Fig. 4. Functional block diagram of crossbar tandem with the first version of CAMA

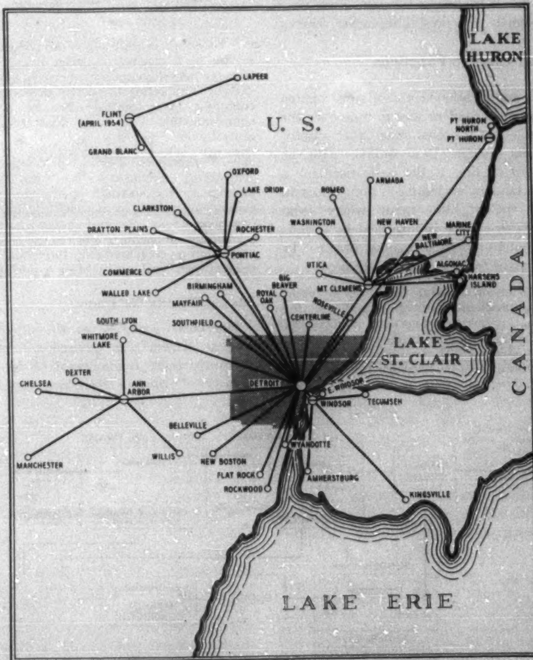

LAKE HURON
U. S.
CANADA
LAKE ST. CLAIR
LAKE ERIE
LAPEER
FLINT (APRIL 1954)
GRAND BLANC
PT HURON NORTH
PT HURON
OXFORD
LAKE ORION
ROMEO
ARMADA
CLARKSTON
WASHINGTON
NEW HAVEN
MARINE CITY
NEW BALTIMORE
DRAYTON PLAINS
ROCHESTER
PONTIAC
ALGONAC
HARSENS ISLAND
UTICA
COMMERCE
MT CLEMENS
BIG BEAVER
BIRMINGHAM
WALLED LAKE
ROYAL OAK
MAYFAIR
ROSEVILLE
CENTERLINE
SOUTH LYON
SOUTHFIELD
WHITMORE LAKE
DEXTER
CHELSEA
ANN ARBOR
DETROIT
E. WINDSOR
WINDSOR
TECUMSEH
BELLEVILLE
WILLIS
NEW BOSTON
WYANDOTTE
MANCHESTER
FLAT ROCK
ROCKWOOD
AMHERSTBURG
KINGSVILLE

Fig. 5. Direct-dialing area via Detroit CAMA at cutover, December 1953

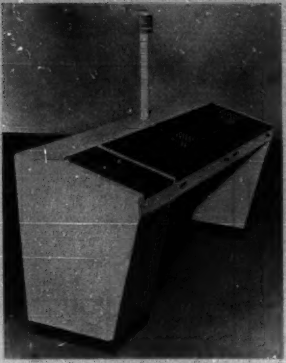


Fig. 6. CAMA switchboard

terms of pulses comprising pairs of five voice frequencies.

3. In the revertive pulsing (RP) language, in which the tandem office pulses its selection progress back to the calling central office with d-c ON-OFF conditions until the called number is matched in the central office.

These four varieties of machine language represent the inpulsing capabilities of the 1956 crossbar tandem system. On the outpulsing side, the system also has full versatility. Its new senders of the PCI, DP, and MF varieties can forward-pulse in any of the aforementioned four machine languages. Need has not arisen thus far, and is not likely to arise, for equipping the RP senders with MF outpulsing. In fact, all the senders can even use a fifth machine language in the forward direction. This is the language of "meaningful silence." It is used, for example, when the tandem office establishes a connection to certain switchboard operators ahead on the straightforward (SFD) basis. In this case, the forward operator receives a request for her action verbally from the originating point through the tandem office.

In view of these machine language capabilities of the crossbar tandem system, it can now be used for interconnections between all the central offices situated within dial cities, between these offices and those in surrounding exchanges, and between different offices in the surrounding area. Fig. 3 portrays these capabilities.

This figure is a block diagram of sources and destinations of telephone traffic that can be routed through the crossbar tandem system. It also shows the types of machine language that may be received from these sources or sent to the destinations. Incidentally, this figure already indicates that the crossbar tandem has outgrown its original mission of interconnecting only local central offices. It may now be also used to set up connections to and from and between long-distance switching systems. More will be said about this later.

EQUIPMENT CENTRALIZATION

As was previously stated, the third fundamental objective of a local tandem system is to effect over-all service economies by centralizing equipment for 1-way message services, and for equipment needed as normal adjuncts of telephone service. In the first category are such existing popular services as up-to-the-minute local weather reports, time-of-day announcements, etc. The centralization of this type of service in the crossbar tandem system required no significant extension of its features; for these purposes, the tandem's function is the normal one of connecting the calling customer to a destination, which is a simulated operator-type central office generally furnishing recorded announcements.

Centralization of equipment needed as normal adjuncts of telephone service is exemplified by the centralization of charging facilities. Simple charging facilities for 1-unit scoring of a customer's message register are, of course, part of dial central offices. However, control arrangements for zone-message charging in panel-type central offices have been incorporated in the crossbar tandem system right from the start. Zone-message charging involves the operation of the customer's message register one or more times on a single call, depending on the distance involved and the length of the conversation. Although the panel central office circuits could be arranged for this feature, it was more economically accomplished by incorporating the charge-calculating arrangement in the crossbar tandem and operating the customer's message register by signals passed over the trunk to the originating central office.

Zone-message registration did not require separate functional control units in the crossbar tandem. It is accomplished as one of the useful by-products of common control system design. From the called central office code, the marker can readily determine the terminating rate zone. Incoming trunks, in turn, can be grouped into originating rate zones, and this information is also available to the marker. Thus, the marker can readily select the appropriate one of the originating-terminating rate-zone combinations and set the incoming trunk for the corresponding charge rate.

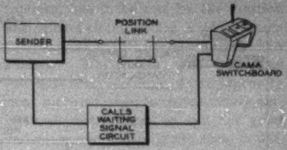


Fig. 7. Association of calls-waiting signal circuit

Obviously, zone registration is satisfactory only for bulk-billed message charging. A far more flexible telephone message billing scheme is the AMA system.[5] In this system the necessary charging information for both bulk-billed and detailed-billed messages is automatically recorded at the central office on paper tape. The record is then processed at the accounting center through a series of information-processing machines.

The introduction of this AMA system was indeed a major step in expanding dial-telephone service with its inherent speed, convenience, accuracy, and uniformity. With AMA recording in a central office, from the standpoint of billing, all restrictions were removed from customer direct-distance dialing (DDD). However, this AMA recording equipment is relatively expensive. In general, its use can be justified economically only in new central offices of the no. 5 and no. 1 crossbar type into which it has been fully integrated. The grafting of this type of automatic recording equipment on thousands of central offices of older design (step-by-step and panel-type) as well as on the many existing No. 1 crossbar offices would be too costly. But the cutomers of these central offices should not be denied the unquestioned advantages of DDD. In this situation, the principle of centralization is applied. By putting the recording equipment at a central point to serve the customers of many central offices, its cost per central office becomes acceptable. Therefore, centralized automatic message accounting (CAMA) has been added to crossbar tandem.[6] With this scheme, if a customer dials a call that requires automatic recording of billing data, the call is routed by the local central office to the crossbar tandem office for such recording.

Unlike zone registration, CAMA required the integration of several major new common control circuits into crossbar tandem. This is shown by Fig. 4.

Fig. 4 is a functional block diagram of

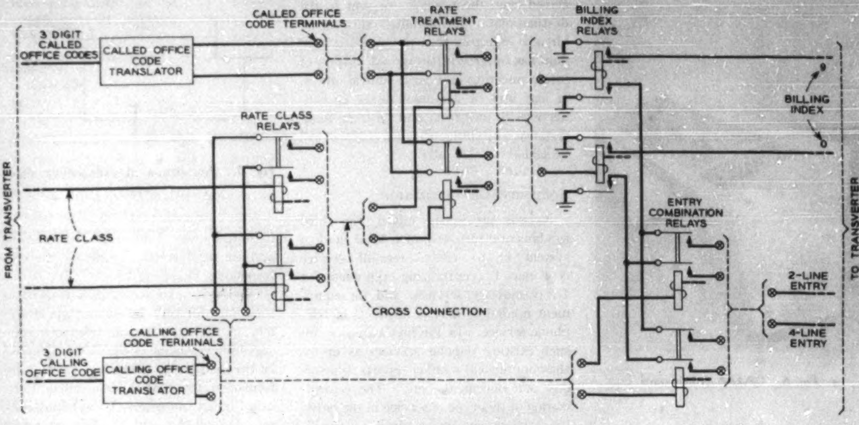


Fig. 8. Simplified schematic of billing indexer

the crossbar tandem system with the first version of CAMA. In this arrangement the sender and marker, of course, retain their previous call-completing functions. However, the sender now has additional major assignments. It must register not only what number is being called but also who will pay for the call, i.e., the calling telephone number. Developments will be completed in the near future to enable the central offices which are working into crossbar tandem CAMA to identify automatically the calling number and pulse this information into the tandem sender over the trunk. At present, to get this information, the tandem sender has to call for the momentary assistance of an idle one of a group of operators. This operator asks the calling customer for his telephone number and, by numerical push buttons, inserts this information into the sender. The sender then passes both the calling and called telephone numbers to the transverter which translates and converts these numbers into the AMA digit-and-word-code structure for subsequent recording. The transverter also requests the billing indexer to determine the applicable charging plan based on the calling and called points and to decide whether the AMA record should or should not include details such as are needed for the customer's toll service statement. Then, under the direct control of the transverter, the recorder perforates into the AMA tape the initial information on the call, including the number of the incoming tandem trunk involved. This last piece of information is furnished by the call-identity indexer. All these CAMA common control elements, including the CAMA operator, finish their respective tasks within a fraction of a second to a few seconds, and are then immediately available for the next call. Meanwhile, the marker has, of course, established the connection, and the sender now completes pulsing forward to the succeeding office. When the called customer answers, the crossbar tandem incoming trunk through the call-identity indexer requests the recorder to punch into the tape the trunk number, and the time of answer, as furnished by the master timer. This same operation is repeated when the customers finish their conversation. These three separately recorded pieces of information on one call are identified by the common trunk number. The CAMA record is processed at the accounting center with the same machines as an AMA record from a local central office.

The significance of this crossbar tandem CAMA scheme can be readily seen. With one stroke, the service advantages of wide-range dialing could now be offered to millions of telephone customers. The first version of this scheme was naturally developed to benefit the greatest number of customers in the large metropolitan areas served mainly by panel and no. 1 crossbar central offices. To expedite early introduction, this initial version was devised so that it requires practically no changes in the central offices it serves. Most of these central offices have had all along the ability (or could acquire it by minor modifications) to make the simple decision whether a call is or is not an extra charge call and, if it is, to transmit it by PCI pulsing to the crossbar tandem office. However, these same central offices have a capacity for 7-digit calls only, and so the initial version of the crossbar tandem CAMA was also a 7-digit PCI system. Even with this limitation, it provided a striking extension of customer dialing. For example, the first installation of crossbar tandem CAMA enabled the customers of 85 central offices in Washington, D.C., and in suburban Virginia as well as Maryland to dial each other directly and to dial their own calls to Baltimore and to other nearby toll points. The second installation enabled 800,000 telephones in 99 Detroit panel and no. 1 crossbar central offices to dial each other directly, and to dial to 63 other communities in Michigan and nearby Canada. The map of Fig. 5 shows this dialing area.

Other large installations of this original 7-digit PCI version of crossbar tandem CAMA are now in service. Some of these subsequent installations, such as the one in New York, opened up a second 7-digit area (New Jersey) for customers to dial into by using the special directing-code of "1-1" for which most of the panel and no. 1 crossbar offices could be readily equipped. This special directing code simply serves to set the central office for transmitting the called number to the crossbar tandem serving the second area

After the introduction of the 7-digit PCI version, development effort has, of course, continued and crossbar tandem now has senders for handling CAMA calls

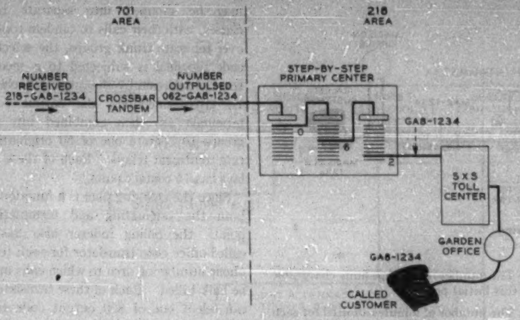


Fig. 9. Code conversion

in the DP as well as in the MF languages; there exists no need for a RP CAMA sender. These latest senders also have a 10-digit-call storage capacity. Thus, the 1956 version of crossbar tandem is designed to handle and record billing data for DDD calls of nationwide scope from central offices that can transmit to it such calls in either the MF or DP language. The former is in anticipation of newly developed storage and outpulsing capabilities that will soon be incorporated into panel and no. 1 crossbar central offices. The latter, in turn, is to extend DDD service to customers served by step-by-step type of central offices, who may dial a simple directing code (such as 1-1-2) to reach the crossbar tandem office and follow, without interruption, with the full called number.

THE CAMA OPERATOR

The transverter, recorder, master timer, and call-identity indexer of the crossbar tandem CAMA system have much in common with the corresponding units of the central office or local AMA (LAMA) system. Hence, they will not be elaborated on in this paper. The two entirely new control elements are the CAMA operator and the CAMA billing indexer. Both of these deserve some remarks.

The over-all function of the CAMA operator has already been stated. In effect, she is a translator, converting the spoken words of the calling customer into digital machine signals acceptable to the CAMA sender, a function which cannot as yet be accomplished economically by electric circuitry. The technical significance of this method of translation is that it poses the problem of integrating a human being with the team of common control units of an otherwise fully automatic system. Such integration, of course, brings its own special requirements.

The first one of these special requirements is that, in contrast to the indifference of the automatic units, a human being needs convenient and pleasant working space. The low, modern, streamlined, sheet-metal, cordless CAMA switchboard of Fig. 6 was designed with this in mind. It provides a convenient work key shelf, and its pleasant color treatment makes a major contribution to a cheerful operating room.

The second special requirement brought about by the integration of operators with the CAMA control circuits is that these control circuits must insure uniform demand for the services of each operator of a working group. An automatic unit does not object to being first choice and thereby to be called upon to carry a heavier work load than a last-choice circuit. In contrast, an operator would. This requirement of highly uniform load distribution has to be met by the position link control circuit of Fig. 4 which hunts for and assigns idle operators to CAMA senders. While this unit comprises standard relay circuitry, much ingenuity went into its design to insure nearly perfect call distribution, in spite of great variations in operator group sizes, in patterns of occupied positions, and sharp fluctuations in incoming CAMA traffic.

Still another aspect of the integration of operators with CAMA comes about because the physical equipment of the crossbar tandem system can be and is often installed in buildings in which there may be no operators' quarters with rest room, etc., facilities. In fact, the location of the building itself may not be the most advantageous from the standpoint of staffing it with operators. Here again, senders, markers, and their likes are not choosy as to location, while operators can be. So, the 1956 version of crossbar tandem offers within-building or remote CAMA switchboard operation; the designed signalling range for remote operation is a maximum of about 30 miles.

These measures are, of course, in the nature of concessions to the human translators. There are, however, factors on the other side of the ledger in connection with this problem of using operators as full partners of automatic control circuits. Control circuits have the same fixed, uniform work time under all conditions; they do not hustle more when the load piles up. In contrast, good operators will increase the tempo of their performance if they receive indication that this is necessary to catch up. For this reason, the crossbar tandem CAMA system includes a calls-waiting signal circuit. Fig. 7 shows how this circuit fits into the system.

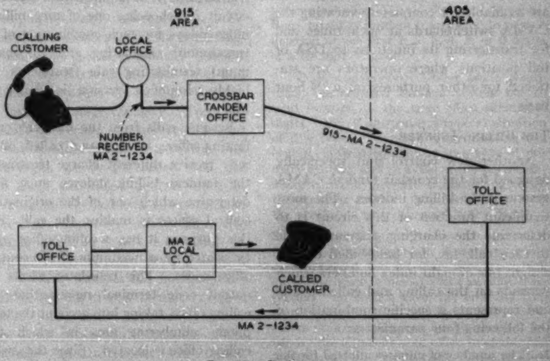


Fig. 10. Prefixing

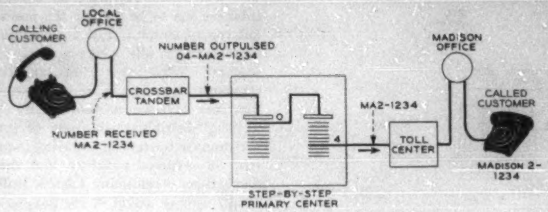


Fig. 11. Prefixing

The lamp standard prominent in Fig. 6 is furnished once in each division of ten switchboard positions and contains three calls-waiting lamps colored green, white, and red. These lamps are controlled by the calls-waiting signal circuit which automatically correlates the number of occupied positions with the number of senders waiting for operators. It can thus indicate the unserved load per occupied position, with increasing load signified by the lighting of the green, white, and red lamps in that order. A multiple of the white lamp also appears on the key shelf of each position. By means of these positional white lamps, the operators know when it is time to put in the extra effort. The colored lamps on the division standard, in turn, enable the supervising operator to regulate the operating team size to match the traffic load conditions.

Another advantage is the flexibility of the human being for doing various tasks in contrast to the highly specialized common control robots. These are used during night hours when CAMA traffic in the crossbar tandem reduces to a negligible level. At such times the robots continue their normal specialized tasks for the occasional call. In contrast, arrangements are available for completely vacating the CAMA switchboards at such times and for transferring its functions to DSA or toll positions where operators are stationed for other purposes on a 24-hour basis.

THE BILLING INDEXER

Another new control unit specifically designed for the crossbar tandem CAMA system is the billing indexer. The most significant function of this circuit is to determine the charging plan applicable to each call, i.e., for bulk-billed calls it assigns one of eight index numbers which depends on the calling and called points and represents a specific combination of the following four parameters:

1. The number of minutes allotted for the initial conversation period.
2. The number of message units chargeable for this initial period.
3. The number of minutes allotted for each overtime conversation period.
4. The number of message units chargeable for each overtime period.

For detail-billed calls, the billing indexer assigns a single index number irrespective of origin and designation. A separate billing index is also assigned to nonchargeable test calls.

As was mentioned previously, the second significant function of the billing indexer is to determine whether the CAMA record should or should not include details such as are needed for the customer's toll service statement. In local AMA offices, the message billing index is relatively easy to obtain as it is practically a function of the terminating office only. In crossbar tandem, however, the complexity of obtaining this index is multiplied by the large number of originating central offices, which may also be in different telephone numbering areas, and, of course, by the extensive bulk-billed tandem terminating coverage.

While the billing indexer is an intricate circuit, in principle it is straightforward. Technically, it is a multistage selecting circuit for choosing one of nine billing-index relays for each combination of 50 (maximum) originating and 75 (maximum) terminating rate treatments. A highly simplified version is shown in Fig. 8.

Because calls from the different originating offices to the same called office may receive different charge treatment, the tandem billing indexer must first determine which one of the originating central offices is making the call. For this purpose, it has a calling-office code translator for a maximum of 800 central office codes. This translator selects an output code terminal representing the calling office, taking into account the telephone numbering area in which the calling office is located. Since the customers of some originating central offices may be grouped into separate rate classes, with their calls to tandem routed over separate trunk groups, the selected code terminal is subjected to a second stage of selection by the rate-class relays. By crossconnections, the output code terminals are then combined into 50 groups to operate one of 50 originating rate treatment relays. Each of these relays has 75 contact pairs.

Since the charging plan is a function of both the originating and terminating points, the billing indexer also has a called-office code translator for each telephone numbering area to which calls may be bulk-billed. Each of these translators can select one of 800 output code terminals representing the called central office in its area. By crossconnections, these terminals are combined into 75 groups and wired to the input side of the 75 contact pairs on the rate treatment relays. On the output side, a contact pair of an operated rate treatment relay thus represents a unique combination of an originating and a terminating rate treatment. By crossconnections, these output contacts are combined into groups to operate one of nine billing index relays. Of these, eight furnish indexes for bulk-billed calls, and one furnishes the index for toll service statement calls.

Fig. 8 does not show the simplification of the billing indexer's job if all calls to a whole telephone numbering area are toll service statement calls. The transverter recognizes such areas as not being one of the areas for which bulk-billing indexes have to be determined. For these calls, the transverter signals the billing indexer that the terminating area is completely toll, and the billing indexer then furnishes the toll index. A similar simplified way is used to assign the tenth billing index to nonchargeable test calls.

To determine whether the CAMA record should or should not contain details (for certain reasons even bulk-billed calls may be detail-recorded), the output code terminals of the calling office translator are also grouped by separate crossconnections to operate one of ten record pattern or entry combination relays. Through their contacts, these relays provide ten different patterns of association between billing indexes and detail or nondetail recording.

Toll Switching Functions

The foregoing presented the 1956 version of crossbar tandem as if it were still primarily a local tandem switching system. It has, however, far outgrown this scope. As a result of developments in re-

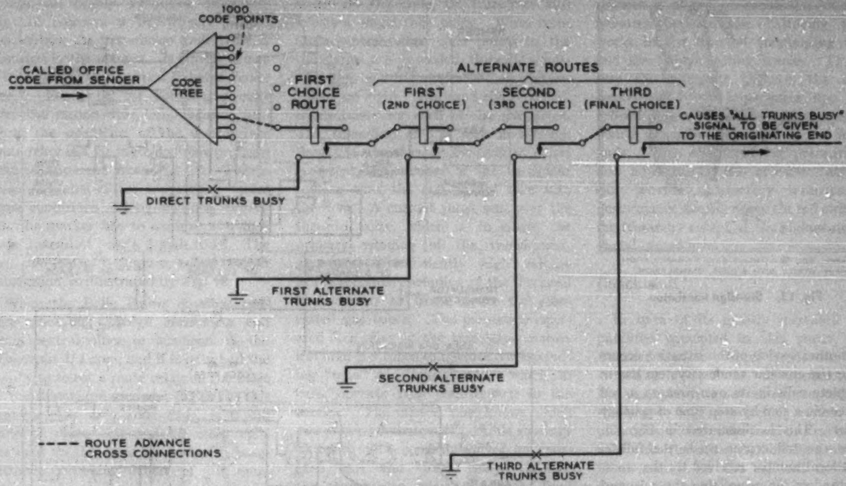

1000 CODE POINTS
CALLED OFFICE CODE FROM SENDER
CODE TREE
FIRST CHOICE ROUTE
ALTERNATE ROUTES
FIRST (2ND CHOICE)
SECOND (3RD CHOICE)
THIRD (FINAL CHOICE)
CAUSES "ALL TRUNKS BUSY" SIGNAL TO BE GIVEN TO THE ORIGINATING END
DIRECT TRUNKS BUSY
FIRST ALTERNATE TRUNKS BUSY
SECOND ALTERNATE TRUNKS BUSY
THIRD ALTERNATE TRUNKS BUSY
---- ROUTE ADVANCE CROSS CONNECTIONS

Fig. 12. Simplified diagram of alternate routing

cent years, it has now also become an important member of the long-distance switching system hierarchy.[7-10] In this hierarchy it can operate at the succeedingly higher levels of toll centers, primary centers, and sectional centers, when subjected to certain transmission requirements which will be discussed in the following. Only the highest level of the long-distance switching hierarchy, that of regional center, is barred to it; this level is reserved for the *4A* toll crossbar switching system which embodies the highest grade through-switching features for long-distance intertoll trunks. The advance of the crossbar tandem to these various levels of automatic switching responsibility is made practical by its basic versatility and is dictated, of course, by the various economies it is able to effect.

The first step in the evolution of crossbar tandem as a toll switching system was reported in 1950.[5] This consisted of the use of crossbar tandem as an office for directly terminating long-distance lines from switchboards in distant cities. Over these lines, distant operators could dial or key-pulse directly into crossbar tandem all calls terminating in local offices to which the tandem had completion access. For instance, the addition of toll terminal equipment at Gotham tandem in New York City in 1947 permitted operators from many states of the country to complete calls to approximately 350 central offices in the New York metropolitan area without the aid of inward operators. This method of establishing telephone connections was a major advance in using crossbar tandem switching equipment for speeding completion of out-of-town calls.

DIGIT-HANDLING CAPABILITIES

To handle outward calls headed into the vast nation-wide telephone network and to perform intertoll through switching, a number of important new features have been developed for the crossbar tandem system. Many of these developments were made concurrently with the CAMA developments, and references to this situation were indicated in the foregoing. For example, it was previously brought out that the latest design dial pulse and multifrequency CAMA senders have a 10-digit call storage capacity. Ten-digit capacity is, of course, the first requirement for nation-wide dialing, where the first three digits specify the called area, the next three the called central office in the area, and the last four the number of the called telephone. Incidentally, this example also brings out the fact that, in contrast to CAMA, the development of toll switching features did not generally result in the use of new common control circuits for crossbar tandem; rather, there was a substantial expansion of the capabilities of the sender and marker circuits.

The second important digit-handling requirement for a fully developed toll switching system is the ability to transmit forward toward the call's destination either the same digits it receives or various modifications of these digits. The modifications take one of three forms:

1. Digit deletion: There are situations where the crossbar tandem system receives the full 10-digit national telephone number, but needs to send forward only part of it. For example, if it has direct intertoll trunks to a common control type of toll center of the called area, it need not send forward the received area code. Or, if it has direct trunks to the central office itself in the called area, it can drop both the area and the central office codes. To meet all conditions, crossbar tandem is arranged to permit deletion of the first 1, 2, 3, 4, 5, or 6 digits of a 10-digit number.

2. Code conversion: Under some conditions, the crossbar tandem system has to send forward to the next office substitute digits. Fig. 9 illustrates a typical situation of this type. To reach a terminating toll center via existing step-by-step type of primary centers, and to avoid costly rearrangement of their present trunking patterns, the crossbar tandem may have to convert the area code into other arbitrary digits.

3. Digit prefixing: It may be necessary for the crossbar tandem system to route calls received for its own area to a toll switching office in the adjacent area for completion. In such a case, the crossbar tandem—being in the terminating area—receives no area code, but has to prefix it for use by the adjacent area office to enable the latter to switch the call back. Fig. 10 illustrates this situation.

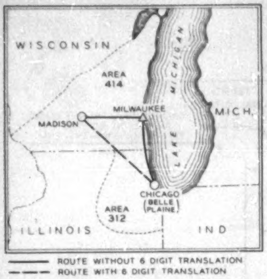

WISCONSIN
AREA 414
MICHIGAN
MICH.
MILWAUKEE
MADISON
LAKE
CHICAGO (BELLE PLAINE)
AREA 312
ILLINOIS
IND
ROUTE WITHOUT 6 DIGIT TRANSLATION
ROUTE WITH 6 DIGIT TRANSLATION

Fig. 13. Six-digit translation

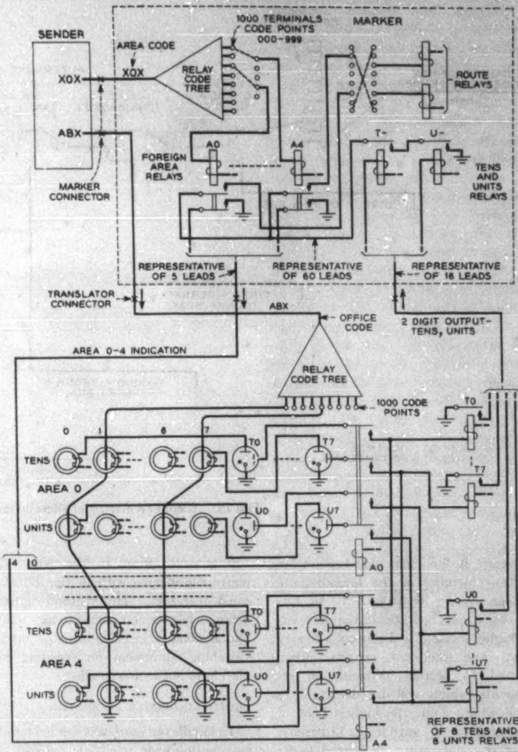

SENDER
XOX
ABX
AREA CODE
XOX
RELAY CODE TREE
1000 TERMINALS CODE POINTS 000-999
MARKER
ROUTE RELAYS
MARKER CONNECTOR
FOREIGN AREA RELAYS
A0
A4
T-
U-
TENS AND UNITS RELAYS
REPRESENTATIVE OF 5 LEADS
REPRESENTATIVE OF 60 LEADS
REPRESENTATIVE OF 18 LEADS
TRANSLATOR CONNECTOR
ABX
OFFICE CODE
2 DIGIT OUTPUT-TENS, UNITS
AREA 0-4 INDICATION
RELAY CODE TREE
1000 CODE POINTS
0
1
6
7
TENS
AREA 0
UNITS
T0
T7
U0
U7
A0
4
0
TENS
AREA 4
UNITS
A4
REPRESENTATIVE OF 8 TENS AND 8 UNITS RELAYS

Fig. 14. Simplified schematic for 6-digit translation

Another version of this situation occurs when the crossbar tandem system has to complete calls in its own area to a toll center via a step-by-step type of primary center. This is illustrated in Fig. 11. Since the toll center needs the full 7-digit local number received by the crossbar tandem for completing the call, and since the step-by-step switches at the primary center use up some digits for routing, the crossbar tandem system must prefix these routing digits to the 7-digit number it receives.

These outpulsing flexibilities reflect new capabilities of both the sender and marker circuits of the crossbar tandem system. The marker's circuitry has been expanded to receive and analyze a great many additional codes and to furnish to the senders additional varieties of instruction for call completion. The sender's circuitry, in turn, has been expanded to execute these new instructions. From the technical standpoint, these new features both in the sender and in the marker have been accomplished by elaborating on existing design techniques rather than by entirely new approaches.

ALTERNATE ROUTING

The same is true of another expanded feature of the crossbar tandem system, namely, the feature of alternate routing. This feature enables the crossbar tandem system to select automatically one of several alternate routes leading toward the destination of a call when the most direct route is busy. It is of special importance in toll switching in that it increases greatly the efficiency of use of the costly long-distance lines with consequent economies. It also provides important service protection when, for some reason, a whole intertoll trunk route fails.

Alternate routing is a natural ability of the marker. As part of its normal operation, it tests the group of outgoing trunks leading most directly to the call's destination prior to making any connection. If all these trunks are busy, it can advance its test to another trunk group in accordance with the wired preference order of its route relays. A limited alternate-routing feature was included even in the earlier version of the crossbar tandem. However, for a high-ranking toll switching system, this original alternate-routing capability had to be expanded, and the 1956 design can choose three alternate routes toward a destination in addition to the most direct route. A highly simplified version of this feature is shown in Fig. 12.

As indicated therein, the called area or central office code received by the marker from the sender results in the selection of a code point which is cross-connected to the first-choice route relay. Each route relay has a number of contacts which are wired to supply all the information to the marker for proper routing of the call. Typical information furnished by the route relay is the type of machine language the sender should use for forward pulsing, the number of digits it should delete, or prefix, or convert, etc. Several of the contacts of these route relays are also used to indicate the number of outgoing trunks to be tested and the location of these outgoing trunks on the crossbar link frames. On the basis of the information from its route relay, the marker tests all of the trunks in the most direct or first-choice route and, if they are all busy, it continues its search for an idle trunk in the first alternate route, the route relay of which is brought into play over the route-advance crossconnection, as shown in Fig. 12.

Like all highly simplified diagrams, Fig. 12 conveys a deceptively simple idea of how the first-choice route relay is operated in the marker. It simply shows a triangle for this purpose labeled "code tree." Especially for completing calls over the nation-wide long-distance network, the operation of the appropriate route relay is a far more complex problem than is apparent from Fig. 12. To operate the first-choice route relay under some conditions of long-distance switching, the marker has to analyze a 6-digit code instead of only a 3-digit code. The need for this new feature, called 6-digit translation, is illustrated by Fig. 13.

When the Belle Plaine crossbar tandem office in Chicago receives a call for a central office in Madison, in the Wisconsin *414* area, and if it only had the ability to select a route relay on the basis of the first three received digits (in this example, the area code *414*), it would have to route this and all other calls destined to this area to the main long-distance switching center of the area, which is the sectional center at Milwaukee. Now, there might be enough telephone traffic from Chicago to Madison to justify direct lines. Routing Madison central office calls over direct lines to the Madison toll office saves trunk mileage as well as eliminates the use of the switching machine at Milwaukee for this traffic. However, to do so requires that the crossbar tandem marker be able to select a first-choice route relay by examining both the 3-digit area code and the 3-digit central office code within the area. Fig. 14 is a simplified schematic of this 6-digit translating feature.

Six-digit translation in crossbar tandem involves the marker and a new control circuit called the "foreign-area translator." This new translator has a capacity for selecting a maximum of 60 separate routings into each of five foreign areas.

On a call, the first three digits received from the sender are used by the marker to select a code point. If a call requires only 3-digit translation, this code point is crossconnected directly to a route relay. If a call requires 6-digit translation, this code point is crossconnected to the proper one of five foreign-area relays. The problem then is to determine which one of a maximum 60 routings into the area should be used to reach the called central office. For this purpose, the foreign-area translator is connected to the marker, and the called-area relay is operated in it. The translator now receives from the sender the second three digits representing the central office code in the called area. By this code, the translator also selects a single code point. Wires from these central office code points in the translator are threaded through transformers which have holes in their centers. There are eight "tens" and eight "units" transformers for each of the five areas. The threading of the wire is through those two transformers for each area that represent the number of the particular routing that the call should take into the area. A current surge sent over the threaded wire, which is, in effect, the primary winding of the transformer, generates a sufficiently high voltage in the secondary winding of the threaded transformers to break down the associated gas tubes. The previously operated area relay in the translator enables the fired gas tubes to operate corresponding "tens" and "units" relays which, in turn, operate their counterparts in the marker. The marker now knows which one of the maximum 60 possible routings to take. The combination of its operated "tens" and "units" relays, through the proper area relay, selects the final code point in the marker which is crossconnected to operate the route relay.

TRANSMISSION ASPECTS

The crossbar tandem is a 2-wire switching system, i.e., it switches only one pair of voice transmission wires over which conversation flows in both directions. Two-wire switching inherently presents more opportunities for objectionable energy reflections or echoes than does 4-wire switching,[11] because of impedance variations that are the result of different calls taking different paths through the office, with varying lengths of cabling involved. The effects of these variations can, however, be minimized to the point where they are not controlling in the assignment of operating losses to the intertoll trunks. This is accomplished by arranging the switching equipment so that the cabling between the various frames of the office is reasonably short and of low resistance, and by building out the capacitance of this cabling to a common value. The balancing networks associated with the 4-wire terminating sets of the intertoll trunks are, in turn, provided with adjustable capacitors for the purpose of balancing this office cabling capacitance.

It is also worth noting that the crossbar tandem system does not include the switching-pad feature which in the past has been widely used in toll switching practice to realize satisfactory singing and echo conditions on each terminating connection. Instead of switching pads, reliance is placed on securing adequate balances between the balancing networks in the intertoll terminating sets and the toll-connecting trunks. To insure this adequate balance, the toll-connecting trunk impedances are improved where necessary, and the short trunks are built out with 2-decibel fixed pads. With the foregoing precautions and adjustments, the crossbar tandem can provide satisfactory transmission performance at all levels of the toll switching hierarchy except at the highest or regional center level.

Conclusion

In spite of its greatly increased capabilities presented in this paper, the crossbar tandem system is not the best switching system for all applications. For example, it is not at all adapted for the largest field of switching system applications, namely, the local central office field. At the other end of the scale, it is no match for the *4A* toll crossbar system in trunk capacity or routing flexibility required at many of the key switching centers of the nation-wide toll network. However, in between these two extremes of the vast Bell System switching system hierarchy, there are a great many positions at the local tandem, toll center, primary center, and sectional center levels for which the crossbar tandem, with all of its new features, is a well-adapted and economical candidate. This is borne out by present Bell System views regarding probable installations during the next decade. According to these views, control switching points, which include all the important long-distance switching points above the toll center level are likely to be of the types shown in Table I.

Table I. Control Switching Points

Type of Switching System	In Service in 1955	Probable for 1965
No. 4 type of toll crossbar	39	58
Crossbar tandem	2	88
No. 5 crossbar	5	53
Step-by-step intertoll	21	59
Other*	1	17
Total	68	275

*Includes types other than Bell and points where type has not yet been determined.

Equally interesting are the present views regarding the probable number of centralized AMA installations by 1965, and the number of main telephone stations that will be served by them, see Table II.

The figures in Tables I and II indicate

Table II. Probable 1965 CAMA Installation

Type of Switching System	Number of Locations	Main Stations Served
No. 4 type of toll crossbar	32	6,800,000
Crossbar tandem	133	21,600,000
No. 5 crossbar	207	3,500,000
Step-by-step	264	7,100,000

convincingly that the crossbar tandem switching system is slated to play a highly important part in rendering good and economical dial telephone service on a nation-wide scale.

References

1. Crossbar Dial Telephone Switching System, **F. J. Scudder, J. N. Reynolds.** *AIEE Transactions (Electrical Engineering)*, vol. 58, May 1939, pp. 179–92.

2. The Number 5 Crossbar Dial Telephone Switching System, **F. A. Korn, James G. Ferguson.** *Ibid.*, pp. 244–54.

3. Fundamentals of the Automatic Telephone Message Accounting System, **John Meszar.** *Ibid.*, pp. 255–69.

4. Automatic Toll Switching Systems, **F. F. Shipley.** *Ibid.*, vol. 71, pt. I, Sept. 1952, pp. 261–69.

5. Crossbar Tandem System, **R. E. Collis.** *Ibid.*, vol. 69, pt. II, 1950, pp. 997–1004.

6. Centralized Automatic Message Accounting System, **G. V. King.** *Bell System Technical Journal*, New York, N. Y., vol. 33, Nov. 1954, pp. 1331–42.

7. Automatic Switching for Nation-wide Telephone Service, **A. B. Clark, H. S. Osborne.** *AIEE Transactions*, vol. 71, pt. I, Sept. 1952, pp. 245–48.

8. Fundamental Plans for Toll Telephone Plant, **J. J. Pilliod.** *Ibid.*, vol. 71, pt. I, Sept. 1952, pp. 248–56.

9. Nation-wide Numbering Plan, **W. H. Nunn.** *Ibid.*, pp. 257–60.

10. Crossbar Tandem as a Long Distance Switching System, **A. O. Adam.** *Bell System Technical Journal*, New York, N. Y., vol. 35, Jan. 1956, pp. 91–108.

11. Crossbar Toll Switching System, **L. G. Abraham, A. J. Vusch, F. F. Shipley.** *AIEE Transactions*, vol. 63, June 1944, pp. 302–09.

The Development of Automatic Manufacturing Facilities for Reed Switches

J. A. HOSFORD
NONMEMBER AIEE

Synopsis: Many unique engineering problems had to be solved before successful automatic manufacturing facilities could be developed to produce high quality reed switches at low cost for the Telephone System. This paper outlines the engineering program that lead to the successful completion of this project.

THE HISTORY of the program to provide high-volume manufacturing facilities for reed switches is of particular interest because it highlights the close association between the final design of a new product and the design of automatic manufacturing facilities to produce it. The design of this switch differs radically from conventional relays and many unique problems had to be solved before successful manufacturing equipment could be developed for this job. From the time of the original conception of the reed switch, it was recognized that widespread application[1] of this improved, general-use switching element would be realized only if it could be manufactured at a very low cost. This objective could be reached only through automatic manufacture. With this in mind, the Bell Telephone Laboratories, in 1950, approached the engineering division of the Western Electric Company on the general problem, outlined their long-range circuit application plans for these switches, and requested that an inter-company program be undertaken to develop the most economical switch design and to find process methods that could be adapted for automatic manufacture. During the following several years, the two engineering organizations worked together very closely. Manufacturing processes were developed that permitted the design of the switch to be reduced to its functional elements. The present-day highly automatic integrated manufacturing machinery for producing the switch is the direct outgrowth of this very successful program.

The dry reed switch as it is known today appears to be about as simple a product as has ever been made. As can be seen in Fig. 1, it consists of an assembly of three parts, two identical reeds and a short piece of glass tubing. The reeds are made of nickel-iron alloy and are gold plated at their contact ends to provide good contact resistance characteristics.

Simple as this product may seem, many unusual engineering problems were encountered during the manufacturing planning stages. Some of these problems were unique and required extensive study before successful solutions were found. Before any new product can find use in the telephone system, it must be capable of being manufactured at a high-quality level and at a cost in line with the older device to be replaced. In the case of the reed switch, it was evaluated against other telephone relays. Basically, the switch performs the same function as the two leaf-spring contacts in the conventional relay and, for this reason, the manufacturing cost goal was established by the known manufacturing cost for two nickel silver leaf springs with their contacts. Upon analysis, this objective turned out to be approximately 1/20 of the cost of other more complex reed switches that were in low-volume production. At this time, the art of manufacturing both mercury and dry reed switches was still relatively new, and the facilities and methods of manufacture were laboratory-life in nature. Switch subassemblies were manually fabricated by loading parts into small jigs, and the success of each operation was dependent to a large degree upon the skill of the operator.

Of equal importance to the low-cost goal, the switch had to be manufactured to very high-quality standards. The electrical characteristics of the switch as determined by its magnetic operate and release measurements had to be held within tight limits, and the entire manufacturing process had to be engineered to assure that all switches would have a reliable long service life. Uniform electrical characteristics required reeds fabricated to very close tolerances, and the machinery and tooling to produce these parts had to be engineered to insure process uniformity. Similarly, the subsequent reed gold-plating and heat-treating operations had to be carefully engineered for process stability.

Unlike conventional relays, the contacts in this switch are not designed to rub or scuff on closure. Contact behavior studies revealed that units with weld-aligned contacts exhibited very long life. Therefore, every effort had to be made to

---

Paper 56-663, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 28, 1956; made available for printing May 8, 1956.

J. A. Hosford is with the Western Electric Company, Inc., Allentown, Pa.

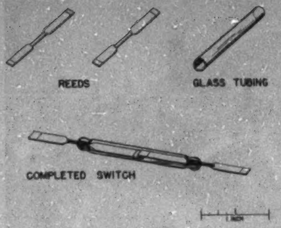


Fig. 1. New reed switch

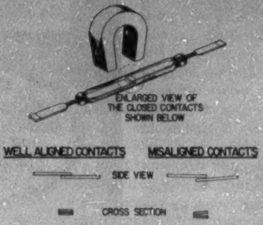


Fig. 2. Contact alignment objectives

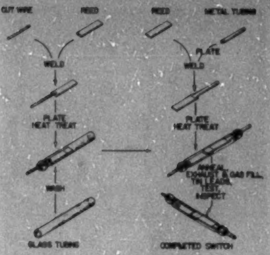


Fig. 3. Old type of reed switch

insure that the reeds in all switches were well aligned, as illustrated by Fig. 2.

To understand better the manufacturing process for glass-enclosed switches, the steps that are followed to assemble one of the older designs of these switches are reviewed. Fig. 3 illustrates one of the typical reed switches that was in limited production at the time this program was undertaken. While this switch is simple in design and contains but five parts, 23 major operations are involved in its manufacture. These include two separate welding operations, four heat-treating operations, three plating operations, two glass-sealing operations employing intricate fixtures, four washing operations, etc. It became obvious that, if the ultimate cost objectives were to be attained, the design of the switch would have to be reduced to its barest elements and a simplified manufacturing process would have to be developed that would lend itself to full mechanization. At the time this project was started, a study was made of all types of automatic machinery employed in the glass-working industry. While several different commercial machines could have been adapted to assemble this type of structure, none were even remotely capable of locating metal elements in glass to the needed accuracy of ±0.0005 inch. It was then realized that it would be necessary to develop a radically new type of assembly machine capable of conforming to the precise needs visualized for the fabrication of switches.

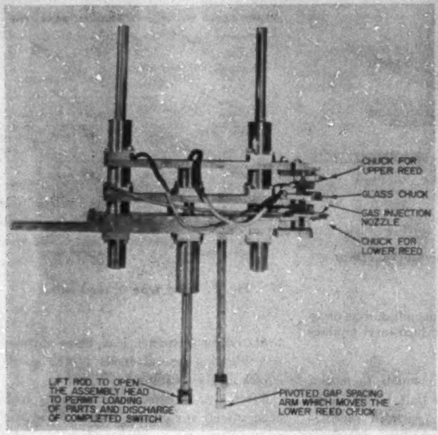
Both the engineering groups involved in this project were aware of these problems and realized that the project would have to be directed accordingly. This program had to be pursued in a logical manner and the work to be done fell into the following four categories:

1. Develop and resolve a suitable manufacturing process.
2. Develop and design automatic manufacturing equipment.
3. Establish final design of the product.
4. Improve and refine manufacturing processes and machinery for high-level production use.

Work was first undertaken to find a manufacturing process that would favor economical production without sacrificing the long life and quality objectives required by the telephone system. After studying all the processes that might be used, detailed cost breakdowns were made for those showing the greatest promise. These studies highlighted the more expensive manufacturing operations and indicated which should be mechanized. One of the more expensive processes revealed by this study was the operation of filling the completed switch with gas. Heretofore switches were gas-filled by one of two methods; neither appeared very attractive for high-level production. The first method employed the bell jar principle where the gas filling was accomplished by performing the glass-sealing operations in a container filled with the desired fill gas. Experience had shown that this process was not well suited to high-level production and had certain inherent process difficulties difficult to correct. The alternative method of gas filling was to use a metal tube as one terminal of the switch. Then after the glass seal was completed, the switch could be placed on a conventional exhaust machine which would cycle to evacuate, gas fill, and weld off the tubing. Inquiries to various suppliers revealed that, even if tubing was purchased in very large quantities, its cost would prevent meeting the switch cost objectives. Thus, it became apparent that an entirely new method of gas filling permitting the use of solid switch terminations would have to be devised. If the air in the switch could be displaced with a suitable fill gas by blowing this gas through the glass tubing during the glass-sealing operations, the need for metal tubing and a welding operation could be eliminated. With the use of such a process, the switch design could be reduced to its barest elements, i. e., three parts—two identical metal reeds and a short length of glass tubing.

After the foregoing assumptions had been made, work was then directed toward developing automatic assembly equipment to determine whether these process assumptions were valid and practical. It was decided that the assembly machine should consist of a number of individual assembly heads mounted on a rotary table. These assembly heads would accurately jig the parts, seal the glass to the reeds, gas-fill the switch, and space the reeds to the desired gap. By building all of the needed accuracy into the assembly heads themselves, the construction of the machine could be simplified and the over-all cost of the machine could be reduced. To eliminate the vibration associated with indexing which, it was felt, was certain to affect the accuracy of the read spacing, it was decided that the machine must have a continuously revolving, rather than indexing turret.

The next step after establishing these general design objectives for the machine was to develop chucks and loading devices which would permit the assembly heads to pick up while in motion all the parts for a switch, i. e., two reeds and a piece of glass tubing. Brass working models of the proposed chucks and loaders were built and mounted on slides to simulate the action of an automatic machine. By working with these models it was possible to modify and improve their design so that they would transfer parts with a high degree of reliability. Work was then started on a complete assembly head. It was decided to design this unit similar to a die set and to build it to very close tolerances to ensure the accuracy of the reed alignment of the completed switch. Fig. 4 pictures the first experimental assembly head. This unit consists of a 3-decked fixture. The top deck holds the chuck for the upper reed. The second deck


CHUCK FOR UPPER REED
GLASS CHUCK
GAS INJECTION NOZZLE
CHUCK FOR LOWER REED
LIFT ROD TO OPEN THE ASSEMBLY HEAD TO PERMIT LOADING OF PARTS AND DISCHARGE OF COMPLETED SWITCH
PIVOTED GAP SPACING ARM WHICH MOVES THE LOWER REED CHUCK

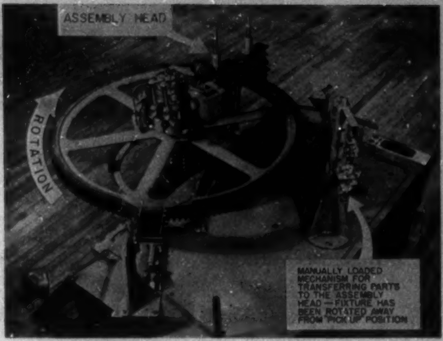

ASSEMBLY HEAD
ROTATION
MANUALLY LOADED MECHANISM FOR TRANSFERRING PARTS TO THE ASSEMBLY HEAD — FIXTURE HAS BEEN ROTATED AWAY FROM "PICK UP" POSITION

Fig. 4 (left). Experimental assembly head

Fig. 5 (above). Experimental assembly machine

holds the glass chuck, and the bottom deck holds the chuck for the lower reed. The two upper decks are lifted vertically by means of a stepped lift rod to permit pickup of the individual switch parts and to insert subsequently the two reeds inside the glass tubing. A pivoted gap-spacing arm located underneath the head causes the lower chuck to move with respect to the rest of the head, thus providing a means for setting the reed gap.

An experimental machine was designed and constructed; see Fig. 5. Parts were manually loaded into a transfer mechanism which swung into position and allowed the head to pick up the two reeds and the glass tubing as it passed by this station; see Fig. 6. Fig. 7 shows a close-up view of the assembly head with the two reeds and glass tubing held in separate chucks. The glass sealing was accomplished with radiant heat from two platinum-alloy coils surrounding both ends of the glass tubing. By passing high current through the coils, a temperature of approximately 1,300 degrees centigrade was attained and sufficient heat was generated to cause the glass to soften and seal around the metal reeds.

By the middle of 1952, work on the experimental machine had progressed to the point where it was possible to assemble a prototype switch. Four different experimental designs of reeds were fabricated and groups of switches were assembled from each of the different types of reeds for statistical evaluation. Mass spectrographic analyses of the residual gas in the completed switches revealed that the gas purity was very high and without a trace of oxygen. Life tests proved that the proposed method for gas filling was satisfactory and could be used in the final machine.

However, checks on the contact alignment in the completed switches revealed that many of the units had poorly aligned contacts; see Fig. 2. Since the reeds used in these experiments had been made by both rolling wire and by coining wire in experimental dies, it was assumed that the alignment difficulty was primarily caused by the crude tooling used to fabricate the reeds. Nevertheless, the analyses of these experimental switches made it possible to freeze the final switch design, Fig. 1, and work was started to design a fully automatic multiple-head assembly machine.

To produce a better reed, a high-speed slide press fitted with a heavy toggle attachment was procured, and production-type dies were built to coin reeds from wire. As soon as this tooling was proved in, switches were fabricated with the improved reeds coined on this machine. These switches still did not exhibit satisfactory contact alignment. A new program was undertaken to improve further the dimensional accuracy of the reeds through better tooling and parts processing. The dies were redesigned with heavier backing plates, improved camming, and carbide punches to withstand

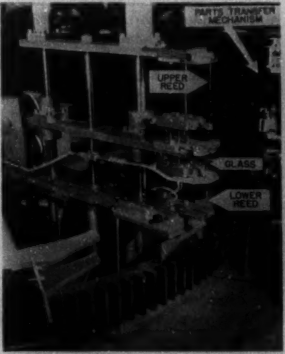

PARTS TRANSFER MECHANISM
UPPER REED
GLASS
LOWER REED

Fig. 6. Experimental assembly head just after picking up parts

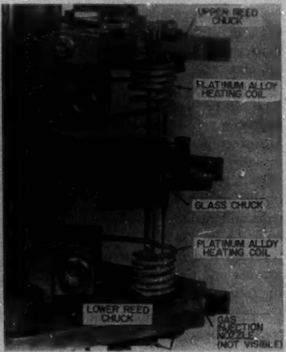

UPPER REED CHUCK
PLATINUM ALLOY HEATING COIL
GLASS CHUCK
PLATINUM ALLOY HEATING COIL
LOWER REED CHUCK
GAS INJECTION NOZZLE (NOT VISIBLE)

Fig. 7. Close-up view of assembly head

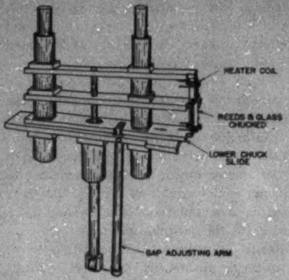


Fig. 8. Schematic of assembly head

the tremendous coining pressures. These new tools produced very straight and extremely accurate parts but the switches still did not exhibit good contact alignment. The general alignment objectives were, therefore, restudied and, in late 1953, a new sequence for assembling the switch was devised that promised great improvement. Rather than hold the upper reed by its chuck during the glass-sealing operation, it was proposed to close magnetically the two reeds and lease the terminal end of the upper reed by opening its chuck. Because the two reeds would be magnetically locked together in the overlap area, the upper reed could move to mate in perfect alignment with the lower reed. With the upper reed then held in this fashion, the upper glass seal could be made and allowed to cool. Then by horizontally spacing the lower reed away from the upper reed with the use of a parallel motion, the contact alignment in the completed switch could be preserved. To assess this proposal, one of the heads was removed from the recently completed final assembly machine and mounted, together with simulated controls, on a bench. With the use of this head, it was possible to assemble switches by the proposed method without disturbing the complex assembly machine. The results were very promising and action was immediately undertaken to redesign and modify the completed assembly machine to adopt the improved method. Two months later, prove-in runs on the multiple-head machine verified that the alignment problem finally had been corrected, and production of switches was started shortly afterwards.

Perhaps the most interesting of all the development problems encountered during this program was the mechanism for setting the reed gap. The reed switch, like most of the electric components manufactured by the Western Electric Company for the Bell Telephone System, had to be manufactured to specific electrical characteristics. The range for the operate and release requirements which are specified in ampere-turns is narrow, and it was recognized that special steps would have to be taken to ensure that all of the switches made would have characteristics which fall into these narrow bands. As repair operations are not possible on this type of glass-enclosed device, the entire manufacturing process had to be carefully engineered and controlled to make sure that the magnetic sensitivity of the switches would be very uniform and of the correct value.

Fig. 10. Automatic test fixture

Like the experimental machine, the assembly heads in the final machine, Fig. 8, are designed to set the reed gap by moving the lower reed away from the upper reed. The reed spacing is accomplished by pushing the lower end of the gap-spacing arm approximately 1/2 inch, which causes the slide on the lower deck of the head to move approximately 0.010 inch. The chuck which holds the lower reed is fastened to the end of this slide. On the rotary machine the gap-spacing arm is moved when the head travels past a fixed cam. Since the reed spacing is a prime factor in controlling the magnetic sensitivity of the

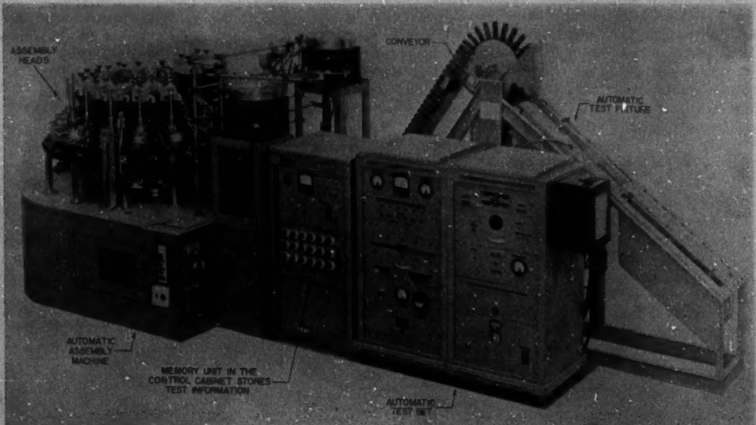

ASSEMBLY HEADS
CONVEYOR
AUTOMATIC TEST FIXTURE
AUTOMATIC ASSEMBLY MACHINE
MEMORY UNIT IN THE CONTROL CABINET STORES TEST INFORMATION
AUTOMATIC TEST SET

Fig. 9. Automatic assembly machine and associated automatic test set

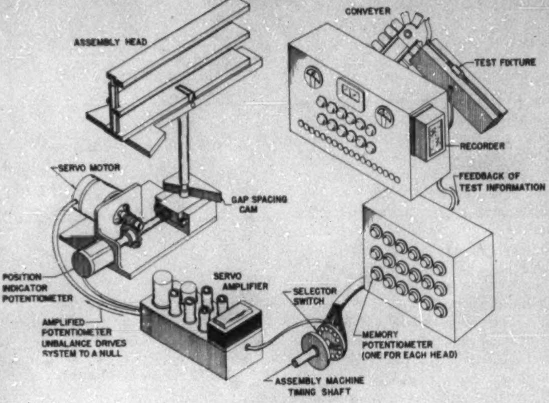

ASSEMBLY HEAD
CONVEYER
TEST FIXTURE
RECORDER
FEEDBACK OF TEST INFORMATION
SERVO MOTOR
GAP SPACING CAM
POSITION INDICATOR POTENTIOMETER
SERVO AMPLIFIER
SELECTOR SWITCH
AMPLIFIED POTENTIOMETER UNBALANCE DRIVES SYSTEM TO A NULL
MEMORY POTENTIOMETER (ONE FOR EACH HEAD)
ASSEMBLY MACHINE TIMING SHAFT

Fig. 11. Schematic of feedback system

switch, considerable attention was devoted to this operation. The mechanical details of the linkage in the assembly head were carefully designed and constructed to exacting tolerances. During the prove-in of the multiple-head machine, it was found that minor variations existed between heads and that other machine factors also influenced the reed spacing in the sealed switches. To compensate for these variations, a servo system was then developed to adjust the gap-spacing cam to a different setting for each of the 18 assembly heads. To understand just how this system works, the entire automatic assembly and test equipment now in use on this job will now be reviewed; see Fig. 9. As mentioned before, the assembly machine consists of a clockwise rotating horizontal table on which 18 heads are mounted. The machine and its associated test set work in unison to assemble and test the switches automatically, and they require only part-time operator attendance to fill the supply line and to correct any troubles.

The assembly operation starts when the heads are fully expanded. The reeds are oriented and fed into the assembly machine by two vibratory feed bowls. The glass tubing is fed from a magazine. Transfer heads, not visible in the figure, deliver the parts to the expanded heads as they pass by these loading stations. As the heads close, the glass and reeds are telescoped together and current is applied to the heater coil at the upper end of the glass tubing. When the upper glass seal is completed, the reeds are spaced apart by means of the gap-spacing cam and the glass is then sealed around the lower reed. After the switch has been allowed to cool, the head is expanded and the switch is discharged into a bucket conveyor.

The conveyor, Fig. 9, is geared to the assembly machine and has buckets associated with each assembly head. In this manner, the identity of the switch is carried over to the automatic test fixtures; see Fig. 10. Here, in addition to 100-percent checks on the operate, release, contact resistance, and voltage breakdown characteristics, a test is made to determine whether the operate value of the switch lies above or below the specification nominal, i.e., 90 ampere-turns. This last electrical information is then fed into a memory device, Fig. 11, in the assembly machine control cabinet and is stored in the corresponding one of 18 potentiometers. At the proper time, this stored information is fed back to a servomotor on the machine. This motor adjusts the sliding cam surface which moves the gap-setting arm on the associated assembly head. The servo system functions by comparing the memory potentiometer setting with the position indicator potentiometer on the sliding cam mechanism. The voltage difference between these two potentiometers is amplified and fed to the reversible d-c servomotor, which then drives the system to a null and effects an exact cam setting as governed by the memory potentiometer.

When a switch under test has an operate value below the specification nominal, the reeds are too close together and a mechanism inside the control cabinet automatically adjusts the associated memory potentiometer upwards a small amount to increase the reed spacing on the next switch assembled on that head. Likewise, if the switch under test had an operate value above nominal, its potentiometer is automatically adjusted downward to decrease the reed spacing on the next switch assembled on that head. Thus, corrective test information is continuously fed back into the assembly machine through the memory device in the control cabinet. The system has purposely been designed to sense, interpret, and feed back information for every switch tested. Experience has shown that frequent machine adjustments are necessary because the platinum-alloy heater coils used to seal the glass to the reeds distort and sag at their elevated operating temperature. When these coils shift in position with respect to the glass tubing, the physical shape of the glass seals will change, and the internal stresses which set up in the glass as it cools will cause a slight shift in the reed gap spacing. The feedback system built into the machine senses the resulting change in the magnetic operate characteristics of the switch and adjusts the reed gap in a compensating direction. Since machined adjustments are continuously made as required, all the switches assembled center around the nominal operate test specification.

This development program was pursued over a 5-year period and has provided truly automated facilities that are now annually producing millions of highly reliable glass-enclosed switches for use in the telephone system at 1/20 of their former cost. The results of this project serve to show how research and manufacturing engineering organizations can work together closely to nurse a high-quality product into low-cost volume production.

Reference

1. Development of Reed Switches and Relays, O. M. Hovgaard, G. E. Perrault. *Bell System Technical Journal*, New York, N. Y., vol. 34, Mar. 1955, pp. 309–32.

Vol. 7
Pt. 1
1956

The Use of Computing Machines in the Statistical Evaluation of Electrical Components

W. E. ANDRUS, JR.
NONMEMBER AIEE

DURING the last few years, the field of applied statistics has experienced a tremendous growth. This is evidenced by the increasing acceptance of the statistician as a vital member of the research and engineering team. The advent of the computing machine is not the least of the factors which has contributed to this growth. Not too long ago, complex statistical problems in multivariate analysis or analysis of variance were beyond the realm of possibility because of the months and years necessary to obtain solutions by conventional hand computation methods. Now, with the existence of computing machines of increased sizes, speeds, and capacities, more of these problems can be solved and are being solved on a routine basis.

Statistics

Applied statistics is a useful tool to the engineer in a number of ways. Statistical evaluations of life characteristics and operating parameters of, for example, tubes, transistors, and rectifiers make the selection of these components and their respective operating limits more precise. This obviously is an aid in the effort to build systems which are more accurate in their function and more reliable in their operation. There is, and always will be, some variation in the characteristics of components and even in the output of a system. If the system variation is excessive, statistical evaluations can determine the sources and differentiate between the contributions of the various components, isolating those sources which have the greatest effect on the output function of the system.

Fundamental Statistical Calculations

The typical statistical problem involves extensive use of rather elementary arithmetical operations. Initially, these operations may be the sums of the data elements or may be the sums of the squares, or specific products of the elements. The sums involved depend to a large extent upon the analysis applicable to the problem of hand. This is usually the first phase of the problem, and once they have been obtained they are used to compute such statistical parameters as the mean, variance, or correlation coefficients.

Take as an example a simple problem in evaluating the pickup time of relays measured at a specific voltage. If a group or a sample of relays is selected at random from a typical production lot and the pickup time of each is measured, a set of data is generated similar to that shown in the following.

Pickup Time, Milliseconds

4.6	4.6	4.5	4.4	4.1
4.5	4.4	4.6	4.9	4.5
4.5	4.8	4.6	4.5	4.6
4.2	4.7	4.3	4.3	4.7
4.8	4.4	4.7	4.5	4.4

$N = 25$ $\bar{x} = 4.52$
$\Sigma x = 113.1$ $\sigma^2 = 0.0361$
$\Sigma x^2 = 512.53$ $\sigma^2 = 0.19$

The well-known average, or mean $\bar{x}$, pickup time is expressed as

$$\bar{x} = \frac{\Sigma x}{N} \tag{1}$$

where x is an individual measurement and N is the number of relays contained in the sample.

The mean, which is perhaps the most widely used statistical measure, is but one of the required values which are necessary to adequately describe a phenomenon. The second, and too frequently overlooked measure, is the standard deviation σ which can be expressed as

$$\sigma = \sqrt{\frac{1}{N-1}\left[\Sigma x^2 - \frac{(\Sigma x)^2}{N}\right]} \tag{2}$$

where σ is the unbiased estimate of the population standard deviation based upon a sample of size N and is a measure of the variation of the data elements about the mean. The two measures then, $\bar{x}$ and σ, are the parameters of the normal distribution which is one of the most important frequency distributions in statistics, since many statistical procedures are based upon it. The use of the normal distribution curve is the same as that of any other distribution curve where the relative frequency with which a variable will take on values between two points is the area under the curve between the two points on the horizontal axis. The term "normal distribution" in this case refers to the fact that the area under the curve is distributed in a specified manner.

Consider Fig. 1, which is a histogram of the data in the foregoing table. The height of the shaded columns shows the frequencies or number of measurements which had a specific value of pickup time x. The smooth curve drawn through the histogram is that of a normal curve and it is evident in this case that the curve approximates the distribution quite well. Fig. 2 denotes some of the properties of the normal curve consistent with statistical theory. As an example, the area under the curve between $\bar{x} \pm 1\ \sigma$ is 67% (per cent). Thus the probability is 67% that a relay selected at random will have a pickup time between $\bar{x} - 1\ \sigma$ (4.33 milliseconds) and $\bar{x} + 1\ \sigma$ (4.71 milliseconds) or that, on the average, 67% of the relays measured will fall within this interval. Correspondingly, the area between $\bar{x} \pm 3\ \sigma$ is 99.7% and, on the average, only three times in a thousand will relays of this type have a pickup time greater than $\bar{x} + 3\ \sigma$ (5.09 milliseconds), or less than $\bar{x} - 3\ \sigma$ (3.95 milliseconds).

Frequently, a determination of the accuracy of the sample mean as an estimate of the population mean is dewired. In this case the population is the total production of this type of relay. Sample means are distributed about the population mean (μ) normally with a standard deviation ($\sigma_{\bar{x}}$) determined by

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \tag{3}$$

where $\sigma_{\bar{x}}$ is the estimated standard deviation of sample means and σ is the population standard deviation estimated from the sample by equation 2. For the example of relay pickup times, $\sigma_{\bar{x}} = 0.038$ millisecond, and there is a probability of 95% that the sample mean $\bar{x} = 4.52$ does not differ from the population mean by more than ±0.076 millisecond.

Statistical Programs

The foregoing example of relay pickup times is not only representative but also serves to illustrate computing techniques and statistical program organization. Note that three basic quantities were re-

---

Paper 56-679, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25-29, 1956. Manuscript submitted March 27, 1956; made available for printing May 18, 1956.

W. E. ANDRUS, JR., is with the International Business Machines Corporation, Endicott, N. Y.

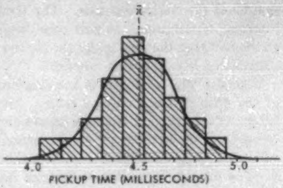


Fig. 1. Histogram of relay pickup times

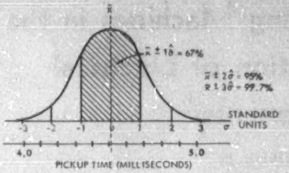


Fig. 2. Properties of normal distribution curve

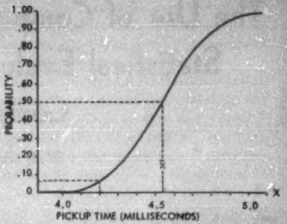


Fig. 3. Cumulative frequency distribution curve of relay pickup times

quired from the data, namely N, $\Sigma\, x$, and $\Sigma\, x^2$. These are the summations mentioned previously and once they are obtained, the original data are not used again for the remainder of the analysis.

The solution of this type of problem then may be separated into two phases. The first phase, in which the required summations are accumulated, is designated as "data reduction." The second, in which the summations are used in the solution of certain equations, is designated "analysis." The computer programs to solve these problems are written in such a way that it is possible to separate the two sections when desired. This is particularly true if the programs are written for a "library" in which they are kept on file and used for subsequent problems.

The discussion of statistical programs will be limited primarily to the IBM (International Business Machine) type-*650* magnetic-drum data-processing machine; see the Appendix. However, mention will be made of another larger computer, the IBM type *701*, which is used to advantage for specific problems of a larger scope.

Return now to the example of relay pickup times. There are many ways in which the data reduction section of this problem may be organized, but for this discussion assume that the data are recorded in punched cards, one data element x per card. Accordingly, the program requirements are that on each card read into the computer the value x is added first to a running sum $\Sigma\, x$, and then squared and added to a running sum $\Sigma\, x^2$. Additionally, the number of cards is counted, and thus the number of data elements, by adding a 1 to the count for N. After all of the data cards are read into the computer the sums are either punched into an output card or stored in memory for subsequent reference.

Since a particular problem might involve a number of groups of data, each group is identified by a problem code and group number, by inserting behind each group of data cards a "distribution control card." When this control card is detected, the summations and identifying designations are punched into an output card. The memory locations containing the summations are then restored to zero to prepare for the accumulations of the next group of data. The output card containing the summations is read into the computer with the analysis program and the parameters $\bar{x}$, σ, are computed and punched into another output card. There obviously are many ways in which programs of this nature may be organized. One significant variation could be to perform the desired analysis after each distribution control break. In doing so, the reading of the summation cards back into the computer would be eliminated.

DISTRIBUTION ANALYSIS

The statistical calculations discussed thus far are usually classified as "distribution analysis." Problems of this nature involve the description of a phenomenon by frequency distributions and polygons, and the estimation of certain characteristics of the population. This is perhaps the most basic type of statistical problem but it is the foundation upon which more complex analyses are based. Thus far only normal distributions have been considered. Since many of the distributions encountered cannot be approximated by the normal distribution, it is advisable to incorporate a normalcy test into the analysis program. One method of performing this test is to compute the index of skewness α_3 a measure of the symmetry with respect to the vertical axis through the mean, and the index of kurtosis α_4 describing the "peakedness" of the curve. For normal distributions, α_3 and α_4 should be very nearly zero. A second method would be to perform a "chi square goodness-of-fit" test in which the observed relative frequencies are compared with the expected relative frequencies for significant deviation. The indexes of skewness and kurtosis are easily obtained by additional program steps in the data reduction and analysis programs. The chi square test, in this case, is less convenient, since the data must be stored in the machine or read into the machine twice in order to perform the necessary comparisons.

It is good practice to include procedures which verify the resulting summations. This can easily be done by including another sum in the data reduction program. If $\Sigma\,(x+1)^2$ is accumulated, it is possible to compare the other sums with it by the identity

$$\Sigma(x+1)^2 = \Sigma x^2 + 2\Sigma x + N \tag{4}$$

or, in the case of the sum necessary for the tests of normality which require Σx^3 and Σx^4

$$\Sigma(x+1)^4 = \Sigma x^4 + 4\Sigma^3 x + 6\Sigma x^2 + 4\Sigma x + N \tag{5}$$

These tests may be incorporated as part of the data reduction or analysis programs.

The distribution analysis programs may take other forms. One which has already been discussed is the determination of the parameters $\bar{x}$, σ, of normal distributions and the indexes of skewness and kurtosis. A further analysis is the cumulative frequency distribution, Fig. 3, which denotes the probability with which observations will fall to the left of any specified value. This type of analysis is particularly valuable to the engineer since it allows freedom as to the selection of probability levels. As an example, 4.2 milliseconds corresponds to a probability of approximately 5%. Thus, 5% of the relays will have a pickup time of less than 4.2 milliseconds and 95% greater than 4.2 milliseconds. The curve is, in effect, the area under the normal distribution curve from minus infinity to the specified point. These areas have been tabled for the normal distribution for values of Z determined by

$$Z = \frac{x_i - \bar{x}}{\sigma} \tag{6}$$

where x_i is the specified point. By using equation 6, x_i in data units is converted to

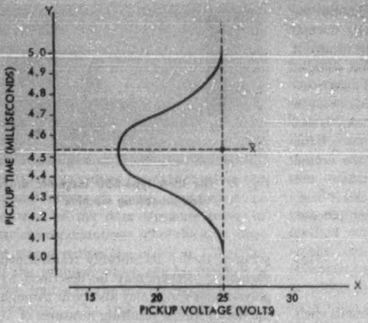
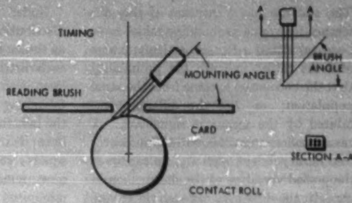


Fig. 4 (left). Distribution of relay pickup times at 25 volts



ANALYSIS OF VARIANCE

DETERMINE WHICH FACTORS HAVE THE MOST EFFECT ON BRUSH BOUNCE

FACTORS ARE:

1. CONTACT ROLL MATERIAL	BRONZE COPPER BERYLLIUM
2. BRUSH MATERIAL	STEEL ALUMINUM COPPER BERYLLIUM
3. NUMBER OF STRANDS	6 8
4. MOUNTING ANGLE	20° 30° 40°
5. BRUSH ANGLE	20° 30° 40°
6. TIMING	

Fig. 6 (right). Variables affecting performance of card-reading brushes

Z, standard units, and one table of areas will serve for all values of $\bar{x}$ and σ. The table of areas may be stored in the computer and with suitable programming the areas determined and punched into output cards for incremental values of x_i.

One of the uses of this type of analysis could be in establishing the performance limits for specific relays. In the example, relays of this type could be restricted to use in circuits which have a timing requirement of 5.2 milliseconds or more. In selecting 5.2 milliseconds it is reasonably assured that only one in 2,000 relays could possibly give trouble when used in this application.

Studies of rectifier characteristics, i.e., forward voltage drop or reverse leakage current, are best accomplished by distribution analysis. Transistors, which are subject to wide variations, also fall into this category. The characteristics of these and other components are best described by their distributions; not only for an accurate description but because, as will be seen in the next section, valid comparisons between manufacturers, operating conditions, or usage are now possible.

Tests of Hypotheses

Problems which are classified under "tests of hypotheses" concern comparisons of two or more items to establish which has the more desirable characteristics. By definition, a statistical hypothesis is a proposed theorem concerning some universe or population of individuals. The process of using a sample to "test" whether the hypothesis is true (or false) is called a statistical proof of the truth (or falsity) of the hypothesis. As an example, the pickup time of one type of relay might be compared with that of another, or rectifiers from one manufacturer compared with those of another, to determine which has the better characteristic. Tests of comparison such as these can best be accomplished by statistical methods.

Consider as an example two groups of rectifiers, each the product of a different manufacturer. The forward voltage drop is measured for each rectifier and the mean and standard deviation of this measurement is determined for each manufacturer. If $\bar{x}_1$ σ_1 are the estimated parameters of the distribution describing the forward voltage drop for the rectifiers from one manufacturer and $\bar{x}_2$, σ_2 for the other, it is possible to determine the significance of the difference between the samyle means. This test, called the "student's test," is expressed as

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sigma_p}\sqrt{\frac{N_1\times N_2}{N_1+N_2}} \tag{7}$$

where

$$\sigma_p=\sqrt{\frac{(N_1-1)\sigma_1^2+(N_2-1)\sigma_2^2}{N_1+N_2-2}} \tag{8}$$

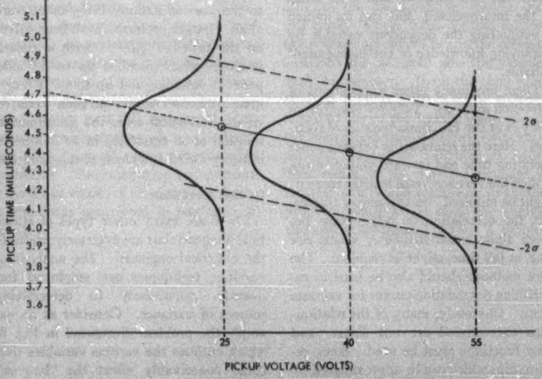


Fig. 5. Linear regression: relay pickup time versus pickup voltage

The statistic t is a measure of this difference for which probability tables have been tabulated. For a particular value of t the tables indicate the probability that the two samples, came from the same population; or that the difference exhibited by the means is significant. A casual examination of equation 7 indicates that t is dependent upon σ_p, a measure of the pooled variation of the data. Consequently, the difference between the means $(\bar{x}_1 - \bar{x}_2)$ may or may not be significant depending upon the uniformity of the measured data and a valid comparison is possible only if the measured characteristic is described by the two parameters, $\bar{x}$, σ.

Another type of test is a comparison of the uniformity of the rectifier characteristics by manufacturer. In this case the test is a comparison of the σ's for each manufacturer and determines whether one manufacturer's product is more uniform than the other. These and other tests are actually the comparison of the distributions, and the input data for computer programs which perform them are the parameters of the distributions. Consequently, programs which handle these problems are the natural extension of distribution analysis programs.

REGRESSION ANALYSIS

Another type of analysis, which is extremely valuable to the engineer and which is handled well by computing equipment is described as "regression analysis." Problems in this area involve the determination of the functional relationships of one variable against one or more others.

Return once again to the distribution of relay pickup times. The distribution was described by a normal curve from which certain information was obtained. Since this was the distribution for relays measured at 25 volts, it can be figuratively rotated 90 degrees as shown in Fig. 4 and a different relationship can be visualized, namely, the effect of voltage. If the pickup time for the same group of relays was measured at 40 volts and again at 55 volts, there would result three distributions as shown in Fig. 5. A straight line drawn through the means shows that there is a linear relationship between the variables, pickup time y and voltage x, and may be expressed as

$$y = a_0 + a_1 x \tag{9}$$

where a_1 is the slope and a_0 is the intercept. Knowing the coefficients of this function it is now possible to predict the mean pickup time for some other voltage, say 35 volts.

Under actual testing conditions the distributions are never perfectly normal as shown in Fig. 5, nor will the means of each distribution fall on a perfect straight line. Accordingly, the method used in establishing the best straight line or even the best function if the relationship is not linear, is that of least squares in which the squares of the deviations of the actual measurements from the function are minimized. In the linear case this is basically the solution of the simultaneous equations

$$a_0 N + a_1 \Sigma x = \Sigma y \tag{10}$$

$$a_0 \Sigma x + a_1 \Sigma x^2 = \Sigma xy$$

The computing load for problems such as these can be quite heavy for any appreciable amount of data, and the computing machine may be used to good advantage. A problem consisting of 100 measurements of (x, y) can be completely solved on the type *650* in approximately 40 seconds. The programming is again broken into the data reduction and analysis phases for flexibility. The input data are recorded, an (x, y) pair to a card, where x is the voltage and y the pickup time. The data reduction program accumulates the sums required in equation 10 and punches them into an output card. When the output card is re-entered into the machine, suitable analysis programs will solve the simultaneous equations 10, for a_0 and a_1, determine the correlation coefficient r, which may be expressed by

$$r = \frac{N\Sigma xy - \Sigma x \Sigma y}{\sqrt{[N\Sigma x - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}} \tag{11}$$

and compute various other values which are indicative of the relationship between the variables. The correlation coefficient r measures the degree of association between the two variables and is an index of the improvement that can be gained in estimating the dependent variable by employing knowledge of the independent variable.

There are many other applications of regression analysis. Take for example an analysis of the life characteristics of rectifiers. Here the relationship between pulse operating time and forward voltage drop, or shelf life versus reverse leakage current, might be required. In the case of transistors the correlation or relationship between alpha cutoff frequency versus rise time, or fall time, might be required. The same methods should also be used in establishing degradation curves for vacuum tubes. Obviously, many of the relationships encountered are not linear, and other functions must be used. Some relationships which can be approximated by exponential functions, may use linear regression theory on suitably transformed variables. Others may be described by polynomials which may also be fit through the data with the resulting measures of fit and dispersion.

Fig. 7. The IBM type-650 magnetic drum data processing machine

A further extension of regression theory concerns problems which involve more than one independent variable. These are multiple-regression problems, which have many applications. As an example, a third variable, namely pre-energization time, could be added to the example of relays. Here a voltage is applied across the relay prior to the actual pickup voltage but of low enough value so as not to pick up the relay The analysis now determines the relationships of pickup time versus various levels of pickup voltage and pre-energization times, and the describing function is a plane, or a surface For the linear case the describing function is expressed as

$$y = a_0 + a_1 x_1 + a_2 x_2$$

where the coefficients are determined by solving the simultaneous equations

$$a_0 N + a_1 \Sigma x_1 + a_2 \Sigma x_2 = \Sigma y$$
$$a_0 \Sigma x_1 + a_1 \Sigma x_1^2 + c_2 \Sigma x_1 x_2 = \Sigma x_1 y$$
$$a_0 \Sigma x_2 + a_1 \Sigma x_1 x_2 + a_2 \Sigma x_2^2 = \Sigma x_2 y$$

This type of problem can be extended to tremendous size involving many variables. Larger problems have been solved on the type-*701 EDPM* with a general program which handles up to 50 independent variables and up to 1,022 5-decimal digit observations of each variable. In the maximum case, the simultaneous solution of 50 equations in 50 unknowns is incorporated as a phase of this program.

OTHER ANALYSES

There are many other types of statistical analyses that are extremely useful to the electrical engineer. The analysis of variance techniques are employed frequently, particularly in determining sources of variance. Consider as an example the problem illustrated in Fig. 6, which outlines the various variables that could conceivably affect the "bounce" of card-reading brushes. The question to be answered is which factor or combination of factors has the greatest effect on brush bounce and thus which should be controlled more closely in order to minimize this troublesome characteristic.

A general program for the type *650* has been developed to handle problems of this nature which are complete factorial designs. Since the computing necessary to reduce the data involves the accumulation of sums, and sums of squares, of the data elements taken in various combinations, all of the data must be stored within the machine. The size of the problem which can be handled is thus limited by the storage capacity of the computer. The type *650* can accommodate as much as 16,500 digits of data with the remaining 3,500 digits (350 10-digit numbers) containing the program. As an example, a 5-factor analysis involving five variables at 3, 4, 5, 6, and 7 levels contains 2,520 observations. If the observations are in 4-digit numbers, there are 10,080 digits and the problem is well within the capacity of the machine. The *650* computes all of the necessary sums of squares by summing over the variables 1, 2, 3, 4, and 5 at a time, and punches these values and the corresponding values of n, into an output card. An additional program converts the sums of squares into mean squares and tests, each with the residual to determine which of the main effects and interactions are significant. Another problem solved recently was a 5-factor analysis involving some 3,000 observations which took approximately 7 hours to compute on the type *650*. This compares with a period of many days required for a solution by conventional hand computation methods.

Another type of analysis used quite extensively employs the auto-correlation and spectral-density functions. These functions determine the presence and magnitude of periodicities exhibited by data which might, for example, describe the performance of an electronic system. One particular problem which was computed on the type *701* considered 1,290 discrete measurements of the performance of an aircraft under adverse wind conditions. The computing and printout of the 1,289 values of the auto-correlation function was completed in less than 30 minutes

Statistical Program Library

The use of high-speed computers is justified without question for problems involving either large numbers of data or extensive numerical manipulation. Problems of a smaller nature may also be handled efficiently by the creation and maintenance of a program library. Programming, coding, and "debugging" account for more than half the time and cost of machine computation, and any method which reduces these items increases the usefulness and value of the computing activity. The program library, in which programs are kept on file and used when appropriate, is but one of the procedures which increases this usefulness.

The general program library usually contains programs which solve simultaneous equations, determine the roots of polynomials, and evaluate transcendental functions such as square roots and logarithms. These programs are classified as subroutines since they may be inserted into and made part of the over-all program for the solution of a particular analytical problem. Many of the programs kept on file perform a complete analysis and may contain a number of the subroutines. Programs for the solution of statistical problems form a unique library because a great many of the analyses may be used over and over again on problems ranging from microsecond timing analyses to million-dollar inventory distributions. The availability of a library such as this permits the use of computers economically for the smaller type of problems which could not otherwise justify their use.

The programs of a type-*650* library may be made compatible with one another by the use of program transfer control cards. These cards contain instructions which are stored in specific memory locations and are read into the machine with the program. Upon the completion of each program, the control is transferred to one of the specific locations and the instruction that is present determines the next phase of the computer procedure. This might be to stop the machine or transfer to a punch routine or to another program which would carry the analysis through an additional phase. A problem is prepared for the computer by the proper selection of library programs and program controls which are deemed necessary.

Conclusions

The computing machine has become a necessary tool in the research and development laboratory and has made possible many of the recent technological accomplishments. Mathematical statistics is also a necessary tool and has become a recognized part of machine development at the IBM engineering laboratories. The correct utilization of these and other technical tools leads to a reduction in the drudgery of routine calculations and to improved utilization of engineering and scientific personnel.

Appendix

The IBM type-*650* magnetic-drum data-processing machine is a stored program, modified single-address, decimal computer with a drum capacity of 2,000 10-digit numbers. A typical computer installation is shown in Fig. 7. The central unit contains the drum storage, the control circuitry, and the arithmetic unit. The unit to the rear contains the power supply, and the circuitry which converts from the decimal number system used in input-output to the biquinary system used internally. The input-output unit housing the card reader and card punch is shown in the foreground. The maximum reading rate is 200 cards per minute and the maximum punching rate is 100 cards per minute. Input and output data are read through a buffer storage which enables the type-*650* to read, punch, and calculate simultaneously.

The 2,000-word magnetic drum is composed of 40 bands each with 50 words of storage. Each word of storage, specified by a 4-digit address, can be used to store data or an instruction word. The average access time to any location is 2.4 milliseconds.

Instructions to the computer are 10-digit numbers (XX|XXXX|XXXX) called instruction words. The first two digits specify one of 44 possible arithmetic or logical operations. The next four digits, the data address, specify the location of the data to be processed, and the last four digits, the instruction address, specify the location of the next instruction word to be interpreted by the *650*. The addition of two 10-digit numbers is achieved in 5.2 milliseconds and the multiplication of two 10-digit numbers requires 9 milliseconds.

Magnetic Data Recording Theory: Head Design

ALBERT S. HOAGLAND
ASSOCIATE MEMBER AIEE

AN ANALYSIS of the process of magnetic recording of digital data is presented from which qualitative design concepts are developed and their usefulness demonstrated, both in the evaluation of magnetic head structures and in their design for high-density storage.

This work was initially undertaken in connection with the development of the International Business Machines *350* magnetic-disk storage file where rather stringent requirements necessitated a program directed towards improving the performance of the existing magnetic recording unit. The external constraints imposed on this work essentially limited the design parameters to those concerned with the magnetic head, excluding the gap dimension. The investigation of this subject led to the development of an analytical formulation of the process of magnetic recording which is directly applicable to data recording. The insight derived from the theory and its relative simplicity in application will be demonstrated through its successful employment in guiding head-design development in connection with the International Business Machines *350* magnetic-disk project.

First, a general discussion of magnetic data recording is briefly presented, stressing the form of the read-back wave form and its relation to bit density. Then the basic theoretical development is presented and the significance of these expressions to magnetic head design is stressed. Experimental results and their correlation with the theory are given, followed by a description of the selected magnetic head for the magnetic-disk file and its performance.

Data Recording

Magnetic recording of data involves the storage and processing of binary information, utilizing two states which are capable of being differentiated. For convenience, consider NRZ (nonreturn-to-zero) recording in which the storage surface is continuously saturated during writing, in one direction for a "1" and in the opposite sense for a "0." A particular input is then composed of a succession of alternating step changes in writing current. Reading involves a derivative-type action, as indicated schematically in the following, which illustrates the over-all transfer process from input to output. The surface co-ordinate is x, and $\mathfrak{x}$ represents the variable indicating the position of the specified surface magnetization relative to the magnetic head

$$i(t)—\mathbf{M}(x)—\phi_h(\mathfrak{x})—vNd\phi_h/d\mathfrak{x}=e(\mathfrak{x})=e(vt)$$

where

$\mathbf{M}(x)$ = distribution of magnetization set up in the storage surface
$\phi_h(\mathfrak{x})$ = reading coil flux as a function of surface position
e = open-circuit read-back voltage
N = number of turns on the reading coil
v = surface velocity
t = time

Hence, $e(vt)$ is a pulselike signal for a step change in $i(t)$,[1] and the surface velocity appears as a scaling factor in both time and amplitude for the output signal. An output voltage signal is then associated with each change in the direction of saturation or reversal in writing current.

Consider the output signals for arbitrary input patterns consisting of sequences of alternating step changes in writing current. On read-back, the existing magnetic field, that was caused by the recorded magnetization of the surface, is extremely weak and, hence, the magnetic head will behave very nearly as a linear element. Further, writing definition or the width of a saturation transition region is much less than the corresponding reading resolution, because of the nonlinear surface saturation characteristic. Thus, generally, reading will limit density before the point is reached at which adjacent changes in surface state modify one another. These considerations imply that the principle of superposition may be applied to the over-all input-output transfer process with the use of the characteristic step function output response. The width and wave form of this characteristic response are thus extremely important, and such responses will be shown as functions of x, or distance, as only in this manner are they meaningful. As indicated in the foregoing, the velocity enters only as a scaling factor between distance and time, and only spatial relations are significant in this recording process. It is evident that bit density considerations will follow directly.

Consider a symmetrical voltage pulse response with a base width of λ. The maximum bit density permissible where no mutual interference is to be allowed, i.e., each saturation change is to be resolved independently, is then equal to $1/\lambda$. On the other hand, if the requirement is only that no pattern modulation occur, implying no signal peak variation, then the bit density will have a limit of $2/\lambda$.

A typical response signal is shown in Fig. 1. The base-line character (or wave form) is quite important, as well as pulse widths at various levels above the base line, in density considerations. The overshoots may be broader than the main pulse. Thus, there is no single criterion that can be used in comparing responses.

A general objective for high-density design is to attempt to obtain a spike-like characteristic output signal, i.e., to eliminate base line signal and maximize the rate of approach of the signal to the base line. Considering undesired signal as noise, the overshoots tend to limit the signal-to-noise ratio at lower densities and first cause overlapping of signals as the bit density is increased. However, even with no base line signal, as the bit density is increased and, correspondingly, a spacing reduction between adjacent sampling strobes, errors will arise with an isolated pulse signal because of the gating of strobes on either side of the strobe which is actually associated with this output indication. This limitation is accentuated by pulse responses with gradual base line approaches and can severely limit performance.

The same considerations apply when the basic input wave form is a discrete pulse rather than a step. The characteristic response in this case will be that from a single pulse input, provided that the resultant recorded dipoles are independent. This case will also be treated.

Paper 56-695, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 8, 1956; made available for printing May 1, 1956.

ALBERT S. HOAGLAND is with the University of California, Berkeley, Calif.

Vol. 7
Pt. 1
1956

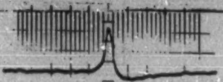


Fig. 1. Typical step function output voltage pulse

Scale: 3.5 mils per division

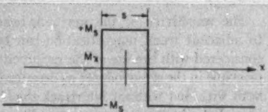


Fig. 2. Idealized recorded dipole

Theory

In addition to the principle of superposition, the principle of reciprocity is used, these forming the basic tools in this analysis. The applicability of the principle of reciprocity to read-back follows as an immediate extension from the justification for utilizing the principle of superposition. As employed here, the concept of reciprocity states that the magnetic field **H** set up by energizing the reading coil with a small current (assuring low field levels appropriate to read-back and, likewise, the same linear behavior) gives a measure at each space point of the degree of coupling existing between a weak magnetic field source at that point and the reading coil.[2] The space points of concern are those that constitute the magnetic sources on read-back, i.e., the storage surface, primarily the track passing under the transducer but also including nearby adjacent tracks.

A reasonable simplification for the objectives desired, in view of the nature of the problem, is to neglect variations with surface layer depth. This is equivalent to considering the surface as a relatively thin film, which is often the case. The depth factor will be treated later. Then functions defined along the surface depend only on one variable. The centerline of the magnetic head gap (ring structures) is, henceforth, chosen as the fixed reference point, i.e., $\bar{x}=0$.

Then

$$\phi_h(\bar{x})=\phi_{hx}(\bar{x})+\phi_{hy}(\bar{x}) \tag{1}$$

where the subscripts denote the reading coil flux contributions arising from each of the two possible components of surface magnetization $\mathbf{M}_x$ and $\mathbf{M}_y$, when a conventional ring-type transducer is used. The y co-ordinate is normal to the surface, defined positive in the direction of the magnetic head. First only a single track will be considered. The fringing field set up along the storage surface track when the reading coil is energized in the manner described for the reciprocity relation will have the vector components $\mathbf{H}_x(\bar{x})$ and $\mathbf{H}_y(\bar{x})$. Then, according to the principle of reciprocity, $\mathbf{H}_x$ is a function giving the sensitivity of the reading coil to horizontal surface magnetization and $\mathbf{H}_y$ gives the reading coil sensitivity to vertical magnetization.

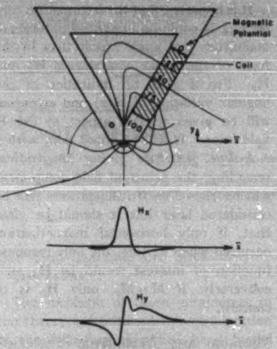


Fig. 3. Field plot showing influence of coil location on H_x and H_y

It can be shown, see Appendix, that

$$\phi_h(\bar{x})=K\int_{-\infty}^{+\infty}\mathbf{M}(x-\bar{x})\mathbf{H}(x)dx \tag{2}$$

where **M** and **H** are vectors combined by a scalar or dot product operation.

Then

$$\phi_{hx}(\bar{x})=K\int_{-\infty}^{+\infty}\mathbf{M}_x(x-\bar{x})\mathbf{H}_x(x)dx \tag{3}$$

and

$$\phi_{hy}(\bar{x})=K\int_{-\infty}^{+\infty}\mathbf{M}_y(x-\bar{x})\mathbf{H}_y(x)dx \tag{4}$$

where the limits of integration are in practice only for the range over which $\mathbf{H}_x$ and $\mathbf{H}_y$ are significant. K is a constant proportional to the surface thickness, if uniform saturation throughout the surface layer is assumed.

Now to clarify the import of these equations, consider an idealized case where $\mathbf{M}_y=0$ and where there is a step change in horizontal magnetization $\mathbf{M}_x$ from $-\mathbf{M}_s$ to $+\mathbf{M}_s$. Then

$$\phi_{hx}=2K\mathbf{M}_s\left[\int_{\bar{x}}^{\infty}\mathbf{H}_x(x)dx+\int_{\bar{x}}^{-\infty}\mathbf{H}_x(x)dx\right] \tag{5}$$

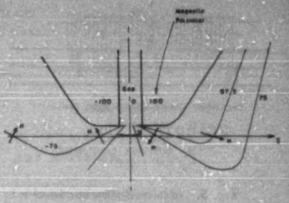


Fig. 4. Influence of narrow pole tips on the weighting function H

and as

$$e(\bar{x})=vNd\phi_{hx}/d\bar{x}=KvN\int_{-\infty}^{+\infty}\mathbf{H}_x(x)\times\frac{\partial\mathbf{M}_x(x-\bar{x})}{\partial\bar{x}}dx \tag{6}$$

then,

$$e(\bar{x})=e(vt)\ \alpha\ v\mathbf{H}_x(\bar{x}) \tag{7}$$

This relation states that the output voltage or pulse response wave form will resemble in time the static field distribution function $\mathbf{H}_x$; i.e., the signal amplitude for a given position of the step change in magnetization relative to the magnetic head gap center line will be proportional to $\mathbf{H}_x$, a weighting function, at that point. Similarly, the output signal arising from a recorded step function change in $\mathbf{M}_y$ would be

$$e_y\ \alpha\ \mathbf{H}_y(\bar{x}) \tag{8}$$

where the subscript on the output voltage indicates that this contribution is due to $\mathbf{M}_y$. Where both components of magnetization exist, the two output signal components may be algebraically added to give the output voltage wave form for $e=e_x+e_y$.

PULSE RECORDING

When discrete pulses are used for data recording, the characteristic voltage response may be related to the sensitivity or weighting functions in the following manner. The magnetization is assumed to be entirely longitudinal, and the following idealized recorded dipole is to be produced by an input current pulse; see Fig. 2.

One direct method for obtaining the characteristic output signal would be to consider merely the recorded pulse as composed of two step function changes in magnetization and to derive the resultant output signal by superposition. However, as will generally be the case, if s is small compared to the width of $\mathbf{H}_x$, then the following approximation may be made, revealing somewhat more clearly the nature of the output signal.

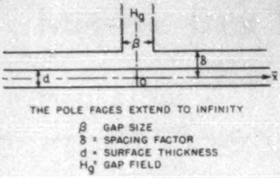


Fig. 5. Idealized magnetic head

The actual $\mathbf{M}_x$ pulse may be replaced by a recorded impulse of the same pulse area. With the use of the relation which exists between the responses to a step function and impulse function in a linear system, the following equations can be written

$$e = e_x$$

and

$$e_x \propto s d\mathbf{H}_x/dx \qquad (9)$$

The characteristic read-back wave form for discrete pulse recording is then seen to be proportional to the derivative of the weighting function and, hence, to the derivative of the characteristic signal obtained when a step function input is used $\mathbf{M}$, being equal to $\mathbf{M}_x$ in both cases.

DETERMINATION OF H

A rapid way to get estimates for the sensitivity or weighting functions $\mathbf{H}_x$ and $\mathbf{H}_y$ is to make rough field plots for the gap fringing field. For this purpose, the magnetic head may be regarded as possessing an infinite permeability. There is no concern about local saturation effects since these maps are to apply for very weak fields. Qualitative information is generally adequate, and normally fine detail would not be justified. It will be seen however that, in general, a factor such as reading coil location is extremely important. The estimates for these fields from such mapping indicate their important qualitative features and these have been confirmed by a powder pattern technique of field observation.

Figs. 3 and 4 show ring-type structures and qualitative determinations of weighting functions, illustrating the importance of coil location and pole-tip shape. The magnetic potential lines shown in Fig. 3 do not, of course, actually apply within the coil or current source region. The orientation of the fringing field vector along the surface, as indicated in Fig. 4, shows the influence of narrowed pole tips in reducing the base line width (neglecting overshoots) of $\mathbf{H}_x$.

It is apparent that the information concerning $\mathbf{H}$ could be presented in terms of magnitude and angle rather than in the form of the scalar components $\mathbf{H}_x$ and $\mathbf{H}_y$. Fig. 4 gives an indication of the angular variation of $\mathbf{H}(x)$ and attention will be given to this aspect of the $\mathbf{H}$ field later. It will be noted that, with a ring-type structure giving longitudinal recording, the principal field component will be $\mathbf{H}_x$. The writing process will be considered later but it should be clear that, if only horizontal magnetization occurred when writing, the only response function of interest would be $\mathbf{H}_x$ and, conversely, if $\mathbf{M}=\mathbf{M}_y$, only $\mathbf{H}_y$ is of concern.

VOLTAGE AND WEIGHTING FUNCTION INTEGRALS

Certain statements may be made about the output voltage-time pulse area, other than its independence of v, in terms of the weighting functions $\mathbf{H}_x$ and $\mathbf{H}_y$. Consider a step change in $\mathbf{M}_x$ from $-\mathbf{M}_s$ to $+\mathbf{M}_s$. Then

$$\phi_h(-\infty) = K\mathbf{M}_s \int_{-\infty}^{+\infty} \mathbf{H}_x(x)dx \qquad (10)$$

and

$$\phi_h(+\infty) = -K\mathbf{M}_s \int_{-\infty}^{+\infty} \mathbf{H}_x(x)dx \qquad (11)$$

Thus, if

$$\Delta\phi_h = \phi_h(-\infty) - \phi_h(+\infty) \qquad (12)$$

$$\Delta\phi_h = 2K\mathbf{M}_s \int_{-\infty}^{+\infty} \mathbf{H}_x(x)dx \qquad (13)$$

but, of course

$$N\Delta\phi_h = \int_{-\infty}^{+\infty} e(t)dt$$

Hence a relation is seen between the output signal voltage-time area and the area under the corresponding weighting function. This immediately gives some insight into the possible degree of overshoot area to be expected. Further, it reinforces the arguments for the importance of reading coil location since, if the coil is so oriented that a uniform horizontal flux field tends to give zero flux linkages, then the negative overshoot area must as a consequence be of the same size as the main pulse area; this applies here to both the weighting function and the output signal which would be proportional.

These same considerations can be as simply extended to an $\mathbf{M}_y$ magnetization component, involving a consideration of the $\mathbf{H}_y$ function.

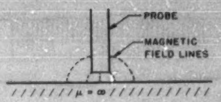


Fig. 6. Field pattern for vertical magnetic probe

ADJACENT TRACK PICKUP

The sensitivity of the magnetic head to adjacent track magnetization can be considered with the use of the reciprocity principle in the same manner as described, both with and without intertrack shielding. Let z be the transverse co-ordinate, and let $z=0$ at the given track and $z=z_1$ at an adjacent track. Note that, since $\mathbf{H}_x(z)=\mathbf{H}_x(-z)$, the worst case for two adjacent tracks, one on either side of the magnetic head, would arise when both were similarly recorded. Now, $\mathbf{H}_x(\bar{x})$ at $z=z_1$ is not only considerably reduced in amplitude over its value at $z=0$ but its relative extent is considerably greater. Since the changes in surface saturation along a track must be alternating, then, as the density of these saturation changes increases, there will tend to be a greater and greater effective cancellation of the individual signal contribution. Thus, because of the low resolution of a magnetic head with respect to neighboring tracks, the lower density patterns are the most troublesome.

FINITE REGION FOR SURFACE MAGNETIZATION CHANGE

An approach to an estimate for the significance of neglecting the actual surface magnetization and using an idealized step function change will be indicated by assuming that $\mathbf{M}=\mathbf{M}_x$ and the change in saturation from $-\mathbf{M}_s$ to $+\mathbf{M}_s$ occurs over the distance x_1 in a uniform manner. Let ρ equal the effective width of $\mathbf{H}_x$. ρ then gives a measure for reading resolution and is equivalent to the factor λ defined earlier for a characteristic output voltage response. Now

$$\mathbf{M}_x = -\mathbf{M}_s + 2\mathbf{M}_s x/x_1$$
$$0 \le x \le x_1 \qquad (14)$$

$$\mathbf{M}_x = -\mathbf{M}_s$$
$$x \le 0$$
$$\mathbf{M}_x = +\mathbf{M}_s$$
$$x \ge x_1 \qquad (15)$$

then

$$\partial\mathbf{M}_x(x-\bar{x})/\partial\bar{x} = -2\mathbf{M}_s(x-\bar{x})/x_1$$

for

$$0 \le (x-\bar{x}) \le x_1$$

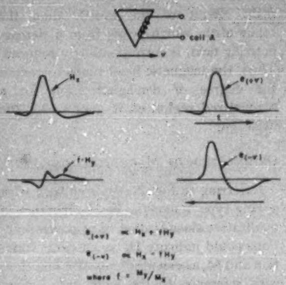


Fig. 7. Effect of oriented surface magnetization on signal

M_y
M_x
+v
M_x
M_y
-v

Fig. 8. Orientation of surface magnetization with direction of motion

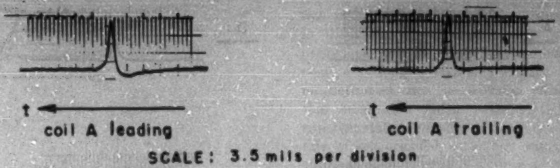


Fig. 9. Wave forms showing effect of direction of surface motion

$$\bar{x} \leq x \leq \bar{x} + x_1$$

and equals 0 otherwise.

Now substituting these conditions into equation 6, the following expression is obtained for $e(\bar{x})$

$$e(\bar{x}) \alpha \frac{\int_{\bar{x}}^{\bar{x}+x_1} \mathbf{H}_x(x)dx}{x_1} \tag{16}$$

which clearly agrees with the earlier expression in the limiting case of a step function magnetization change. The influence of x_1 on the output signal can be quickly estimated. For any value of x_1, $e(\bar{x})$ can be obtained by averaging $\mathbf{H}_x(\bar{x})$ over the interval x_1 for a set of values of $\bar{x}$. This procedure is readily done by graphical means.

The principal influence of a finite x_1 is, of course, to reduce the signal peak and increase the base line pulse width. It is evident that, for an x_1 only several times less than ρ, the approximation of the saturation change by a step function is very good. Experimental attempts to determine a magnitude for such a factor on oxide surfaces indicate that the actual situation is even more favorable to the foregoing approximation. This is in accord with earlier statements regarding the importance of the reading resolution as compared to the writing definition, because of the surface saturation characteristic, since x_1 is a measure of the latter while ρ gives a measure of the reading resolution. For a surface with a "perfect memory"

$x_1 = \rho/2$

and

$\lambda = 3\rho/2.$

Gap Size, Spacing, and Surface Thickness

The inclusion of these parameters in the theory are outlined by consideration of the following simplified head model and with the assumption of a recorded step change in longitudinal magnetization. The extension of this development to other cases is obvious and it particularly lends itself to a ready description of the manner in which gap size, spacing, and surface thickness enter into the determination of the characteristic response. Here, only $\mathbf{H}_x$ need be considered; see Fig. 5.

First neglect the surface layer depth. Then, the following expression may be written

$$\mathbf{H}_x(0) = g(\delta/\beta)\mathbf{H}_g \tag{17}$$

where $g(\delta/\beta)$ is a monotonically decreasing function and $g(0)$ is somewhat less than 1.0. The function g gives the gap fringing field attenuation. The gap size and spacing distance enter in the manner shown for this particular model for here parameter variation actually only involves scaling, a single fringing field plot sufficing. Thus, if both β and δ are changed in the same ratio, this merely amounts to a magnification or reduction of the entire field plot, and the relative field attenuation given by g would remain the same. For an actual head structure, the fringing field dependency upon these parameters can again be estimated fairly quickly by mapping. Now almost the entire magnetomotive force will appear across the gap when the reading coil is energized. Thus, $\mathbf{H}_g\beta = Ni$. Then

$$\mathbf{H}_x(0) \alpha g(\delta/\beta)/\beta \tag{18}$$

The foregoing relation indicates the dependence of the output signal amplitude, for the step function change in $\mathbf{M}_x$, upon the gap size and spacing. If both these factors are halved, then g remains the same and the relative magnitude of the peak value of the weighting function and, hence, the output signals are doubled. It is clear that for this idealized head, these reductions are equivalent to scaling the field plot by a factor of one-half in each direction and, therefore, the relative pulse width will be reduced to one-half its former value. Thus, the weighting function integral value is preserved, as would be expected from the previous discussion on this subject.

The surface thickness may be included in the following manner. Let $\mathbf{H}_x(\bar{x})$ be replaced by $\bar{\mathbf{H}}_x$, the average value of $\mathbf{H}_x$ over the surface depth. It was pointed out that the constant K in the expressions for flux is proportional to the surface thickness, for uniform magnetization throughout the surface layer, since the total magnetic field source coupling with the reading coil for any value of $\bar{x}$ is proportional to d. Likewise, K would be proportional to the head width with the 2-dimensional type of problem here. Then

$$e(0) \alpha d\bar{\mathbf{H}}_x(0) \tag{19}$$

When $\bar{\mathbf{H}}_x(0)$ is a good approximation for the average value of $\mathbf{H}_x$ over the surface layer cross section in the plane $\bar{x} = 0$, the following expression, indicating the signal amplitude dependence upon gap size, spacing, and surface thickness, may be written

$$e(0) \alpha g(\delta/\beta)d/\beta \tag{20}$$

Note that δ is measured to the center of the storage surface layer.

Vertical Probe

A vertical recording probe will be briefly considered to indicate that these concepts are not limited to any particular structure. Consider the rectangular probe shown in Fig. 6 and assume that the storage layer is on the surface of an infinite permeability undercoating. The magnetic field lines produced upon energizing a coil on the probe then have the form indicated, entering the undercoating normally which, for this case, is of infinite permeability.

Now with this form of recording, the surface will be magnetized in the vertical direction and, hence, for a step change

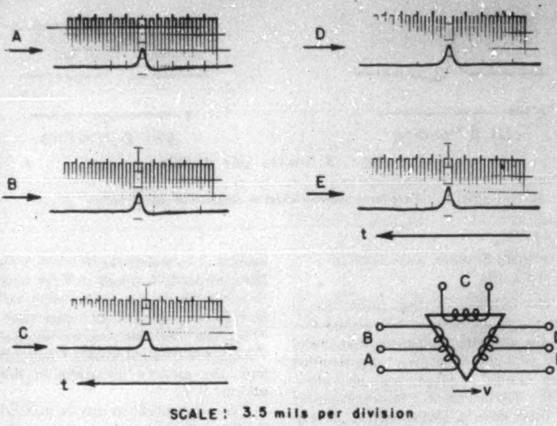


Fig. 10. Wave forms showing influence of reading coil location

in magnetization, the output signal will be proportional to the $\mathbf{H}_y$ weighting function. It is apparent here that $\mathbf{H}_y$ is symmetrical about the probe axis and, further, that $\mathbf{H}_y(\bar{x})$ will be of the same sign for all values of $\bar{x}$ near zero. These facts immediately indicate that a symmetrical output pulse having no overshoot would be expected for such a recorded step change, the return flux diffusing through a sufficiently large return path, removed from the probe region, for an actual device.

SURFACE MAGNETIC CHARACTERISTICS

Writing

The actual width of a recorded steplike change in magnetization, previously designated as x_1, is a function of the rectangularity of the saturation hysteresis loop of the magnetic storage surface material. The nonlinear saturation characteristic of the surface tends to establish a threshold writing field value which must be exceeded to alter significantly the prevailing saturation state during a reversal of magnetization. Consideration of the nature of the actual mechanism by which a small region is reversed in the environment of the opposing saturated volume would indicate a starting field greater than the coercive force by an amount related to the material magnetic properties. These facts limit the effective range of the writing fringing field, this being important on that side of the magnetic head for which the surface will not later pass through the gap region. The more rectangular the surface hysteresis characteristic the narrower will be the transition width x_1.

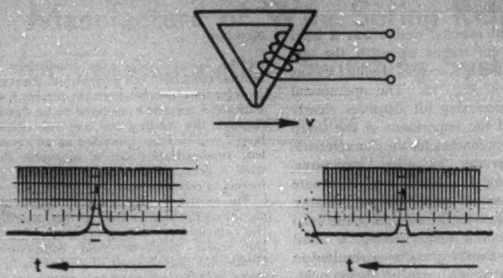
DEMAGNETIZATION

As and after a surface region is magnetized, a stable magnetic state is reached, dependent on the recorded magnetization and its self-demagnetizing field. This situation exists as long as this surface region is removed from the vicinity of the magnetic head. For a sequence of alternating steplike changes in surface magnetization, the demagnetizing field may be visualized as arising from a linear chain of opposing horizontal bar magnets the lengths of which depend on the separation existing between the step changes in magnetization.

Under equilibrium conditions, the surface magnetization will be the same after passing the magnetic head as it was prior to coming into the neighborhood of the magnetic head. However, each time a surface region passes by the magnetic head, there arises the possibility of a cyclic variation in the surface magnetic state. The line integral of the magnetic field intensity about any closed path is equal to zero in the absence of current sources. Consider a path, one branch of which extends along the surface layer center line, linking two surface points located on opposite sides of the magnetic head gap. To the extent that the presence of the magnetic head reduces the contribution to this integral along the external surface path (consider the possible choice of a path following the magnetic circuit of the head), then the other term, which arises from a demagnetizing field, is correspondingly reduced. Thus, the magnetic head tends to reduce the degree of demagnetization for a magnetized region as it goes under the head gap.

ORIENTATION OF MAGNETIZATION

Referring to Fig. 3, it is seen that with a ring-type magnetic head structure a qualitative classification of response functions could indicate $\mathbf{H}_x$ as an even function and $\mathbf{M}_y$ as essentially an odd function, with reference to the head gap. Now consider a step change in saturation magnetization where the direction of saturation is at some angle with respect to the horizontal, less than 90 degrees. Let $\mathbf{M}_x$ be the principal component, a positive step change being assumed, and first consider the case where $\mathbf{M}_y$ is also a positive step change. Now since the actual change in surface magnetization is composed of these two similar terms, the output response is the sum e_x+e_y, each term of which is proportional to its corresponding weighting function, adjusted by the actual magnitude of the associated step change in magnetization. As is clear from Fig. 7, where this composition of signals is shown, the resultant output signal, $e(+v)$ is unbalanced relative to $\mathbf{H}_x$. Now regard $e(-v)$ in Fig. 7, showing the same composition with only the sign of $\mathbf{M}_y$ changed. It is evident that this is a different characteristic output signal possessing another wave form unbalance. The base line wave form of the output signal is considerably different in these two cases. This effect is easily visualized in terms of the angular variation of $\mathbf{H}(\bar{x})$. It follows as a direct extension from the earlier theory that for an oriented step change in magnetization the output signal for each value of $\bar{x}$ is proportional to the scalar product of $\mathbf{H}(\bar{x})$ and a unit vector in the direction of $\mathbf{M}$ or, in other words, proportional to the magnitude of the component of $\mathbf{H}$ in this direction. Now regard in Fig. 4 the region giving rise to the larger $\mathbf{H}_x$ overshoot, this region being located on the coil side (right-hand side) of the magnetic head. It is seen that $\mathbf{H}$ is oriented along an axis extending into the fourth quadrant, obtained by a clockwise rotation of the horizontal axis. Then, for $\mathbf{M}$ oriented along an axis traversing the first quadrant, corresponding to the case giving rise to $e(+v)$ in which both $\mathbf{M}_x$ and $\mathbf{M}_y$ have the same sign, it is evident that the scalar product occurs between two vectors at nearly 90 degrees



SCALE: 3.5 mils per division

S-29 (30 mil radius pole tip) S-25 (2 mil radius pole tip)

Fig. 11. Effect of pole-tip radius

with respect to one another. The change in sign of $\mathbf{M}_y$, used for obtaining $e(-v)$, rotates $\mathbf{M}$ by 90 degrees and, therefore, orients it along an axis running into the fourth quadrant and much more closely parallel to the axis giving the orientation of $\mathbf{H}$ in the region of concern. Thus, it would be expected that the overshoot signal response of $e(-v)$ would be considerably more pronounced.

The significance of these two characteristic wave forms will become apparent if the writing process is briefly considered. The storage surface passes through a 2-dimensional recording field when writing. The important portion of this field, as far as writing is concerned, is the field that the surface passes through on leaving the gap region of the magnetic head. Here a surface cross section horizontally saturated in the gap itself is subject to a not inappreciable vertical magnetizing component as it passes out of the vicinity of the head. Then there will be a tendency for an orientation of the saturated surface magnetization away from the horizontal. Further, this orientation will depend on the direction of motion of the surface during writing. The two possibilities are indicated in Fig. 8. It can be seen that two characteristic output voltage wave forms might be expected, the particular one depending upon the direction of surface motion during writing, except for an entirely symmetrical recording structure. A reversal of the direction of relative motion on reading adds no additional wave forms but merely produces a mirror image of the wave form previously obtained. Fig. 9 illustrates actual characteristic read-back voltage wave forms showing the influence of the direction of surface motion. A delta-type head structure was used with a coil on one leg, as indicated in Fig. 7. The leading and trailing leg designations indicate that the coil leg is the one first passed by a surface point and vice versa respectively. Note that the type of oriented magnetization expected is in agreement with that necessary to account for these wave forms in this theory.

Theory as Applied in Design

A close interaction existed between the experimental work and the refinement and application of the theory. The concepts and relations described readily lend themselves to the interpretation of existing experimental results from which performance predictions can be made for contemplated structures. The concept of an oriented saturation magnetization was essentially developed, as outlined in the foregoing, to explain early experimental results for which an assumption of a longitudinal step change in magnetization was inadequate. While this orientation factor could not be experimentally determined, this hypothesis of an oriented saturation magnetization led to a definite pattern of behavior which could be examined. All the tests indicated consistent agreement with this picture. The procedure for head design was to use only the step function response for evaluation. Final tests were carried out with several structures with the use of recording circuitry and the magnetic head performances under operational conditions correlated excellently with the step function response results.

On the basis of the proportionality between the output signal and the weighting functions $\mathbf{H}_x$ and $\mathbf{H}_y$, the importance of reading coil location and pole-tip shape is readily apparent. Further, with oriented magnetization, even with a symmetrical core structure, the leg upon which a reading coil may be wound is of paramount importance. The prior discussion and Fig. 7 indicate that a favorable location for a leg-mounted reading coil would be the trailing leg with respect to the surface motion since an overshoot compensation is possible.

This design theory indicates quite obviously the desirability of reducing gap size, spacing between head and surface, and surface thickness from the point of view of resolution. In the experimental results presented, these parameters were fixed and, hence, these results show an optimization within this framework.

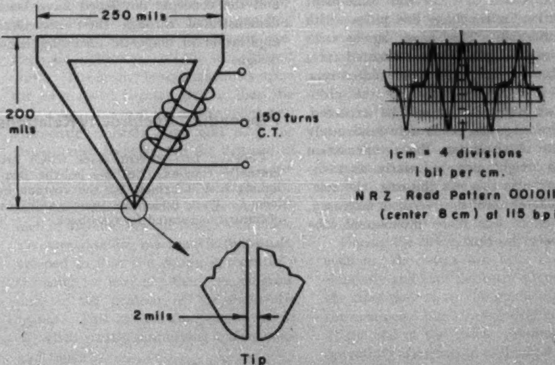


Fig. 12. Selected design and read-back pattern obtained

Experimental Results

The following wave forms were obtained in the course of the head design program for the International Business Machines *350* magnetic-disk project. Some results are from structures constructed primarily to test the usefulness of the conceptual approach described here. Some results then merely illustrate certain points. Following these, the selected head design and its performance will be indicated.

Fig. 10 shows a delta-type head and the signals obtained on read-back from coils located at various positions around the delta for the same step function change in saturation. The relative influence of coil position on overshoot can be noted. An indication of the compensation effect for the trailing leg coil position on overshoot may be seen. The leading and trailing leg identifications apply to the head orientation with respect to surface motion during writing. The signals obtained when a completely different ring-type head structure was used for writing with, however, the same gap size indicated that the recorded magnetization is primarily dependent on the gap field and not greatly influenced by pole piece shape.

Fig. 11 shows the characteristic output signal for two delta structures, both with the favorable trailing leg coil location but one (*S-25*) possessing a pointed-type pole-piece design compared to the other (*S-29*) of which the pole-tip edges adjacent to the surface, are withdrawn much more gradually. The compensation of overshoot is such that, in so far as overshoot is concerned, the responses are nearly equivalent. However, the greatly increased pulse steepness in base line approach of *S-25* and consequent reduction in the base line pulse width is noticeable. This result agrees with that expected from the estimated form of $\mathbf{H}_x$. In one sense this pointed-type head is a tailored design for the problem at hand. As would be expected, if these magnetic heads were deliberately spaced much closer, the compensation action might not be nearly as satisfactory. As this was the case, the relative magnitude of the overshoot increased. This effect was more pronounced, the narrower the tips.

DESIGN RESULTS

The final selected design was a head with a 2-mil radius tip with the coil assembly located on the trailing leg, as indicated in Fig. 12. The operational tests at increasing bit densities clearly established the importance of the sharp base line approaches for the characteristic pulse response, as other structures, while lacking this feature, did eliminate the overshoot as well. The performance of this head under the nominal operational conditions of a 1-mil spacing, a 2-mil gap, and a 1-mil surface layer resulted in a conservative estimate of about a 13-mil base line pulse width for the characteristic response and a peak-signal-to-peak-overshoot ratio of greater than 20. Nominally, no pattern modulation or read-back amplitude attenuation, would be expected below 150 bpi (bits per inch) storage density. Fig. 12 shows a pattern read-back at 115 bpi, a figure which represented an equipment limitation at the time. The specified maximum operating density was 100 bpi.

Conclusions

A theory of magnetic data recording has been presented and developed which is ideally suited to this type of recording. Qualitative principles are made available as guides for design and allow a direct approach to the evaluation of structures and the interpretation of experimental results. Further, these methods can be usefully applied from crude estimates to any desired degree of refinement, and they lend themselves directly to a study of recording structures of a more radical nature. The theory gives considerable insight into the subject of magnetic data recording. The value of this theory and the concepts developed have been demonstrated through their successful application in magnetic recording head design.

Appendix. Reciprocity Relations

Consider two current loops which are mutually coupled. Let ϕ_{ij} be the flux-linking loop i caused by the current in loop j. Then, under the same conditions of linearity as stated for read-back

$$\phi_{12} = L_{12} i_2$$

and

$$\phi_{21} = L_{21} i_1$$

where $L_{21} = L_{12}$. L_{12}, the coefficient of mutual induction, is a geometrical constant. Now for the objectives here, the current i_1 will indicate a current considered to be flowing through the reading coil and the current loop 2 will be later identified as an equivalent representation for a magnetic shell with a permanent magnetic moment, the foregoing relations applying for either case.

The most convenient way to develop the equations for the reading coil flux, based on the principle of reciprocity, is to consider the scalar magnetic field components. Hence, the justification of equations 2, 3, and 4 will be demonstrated by considering the reading coil flux contribution ϕ_{hx} which arises from the horizontal magnetization component $\mathbf{M}_x$ of the surface layer. It will be assumed that no functions vary over the surface layer depth.

Again let $\mathbf{H}_x(\bar{x})$ be the x-directed magnetic field intensity set up along the surface by a unit current through the reading coil. Then, a current i_1 energizing the reading coil will establish at the surface a magnetic field intensity proportional to $\mathbf{H}_x i_1$. The flux traversing a track cross section located at the position $\bar{x}$ with respect to the magnetic head is then proportional to $A\mathbf{H}_x i_1$, where A is the cross-sectional area of the track. L_{12} is thus seen to be proportional to $A\mathbf{H}_x$, according to the previous relation. For a current loop of differential width encircling this cross section and carrying a current i_2, it is immediately evident that the flux linking the reading coil caused by this current can be expressed as follows

$$d\phi_{12} \propto A\mathbf{H}_x i_2$$

Now this current loop can be replaced by an equivalent magnetic shell with a magnetic moment $\mathbf{M}_x$ proportional to Ai_2. Thus, because of this magnetic shell, the reading coil flux which, as it arises from $\mathbf{M}_x$, will now be designated as ϕ_{hx} can be expressed in the following manner

$$d\phi_{hx} \propto \mathbf{M}_x\mathbf{H}_x$$

The surface layer is actually composed of a continuous distribution of magnetization and, thus, the total flux ϕ_{hx} is given by the integral shown in equation 3.

References

1. MAGNETIC DRUM RECORDING OF DIGITAL DATA, Albert S. Hoagland. *AIEE Transactions*, vol. 73, pt. I, Sept. 1954, pp. 381–85.

2. STUDIES ON MAGNETIC RECORDING, PART II, W. K. Westmijze. *Philips Research Reports*, Eindhoven, Holland, vol. 8, June 1953, pp. 161–83.

3. FIELD MEASUREMENTS ON MAGNETIC RECORDING HEADS, D. L. Clark, L. L. Merrill. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 35, Dec. 1947, pp. 1575–79.

Manufacture of Wire Spring Relays for Communication Switching Systems

JAMES W. RICE
NONMEMBER AIEE

FROM THE standpoint of design objectives, features, and operating characteristics, the Bell Telephone System's new wire spring relay for telephone switching systems has been described in a previous paper.[1]

Millions of these relays, Fig. 1, are now being produced annually at the Western Electric Company's Hawthorne plant in Chicago. To manufacture them in this quantity and at low cost, has required the development of machines and techniques which were unknown a few years ago. A description of some of these manufacturing facilities and their operation is the subject of this paper.

As the name suggests, one of the principal features of the design is the use of round wires rather than flat strips to make up the contact spring arrangement. These wires are molded in phenolic plastic blocks which hold the wires in permanent alignment, insulate them from each other, and which are specially shaped to facilitate their assembly into relays. In general, each relay uses three of these molded wire subassemblies. The manufacturing processes and equipment developed for their mass production will be described.

Only two basic types of wire spring assemblies are required, Fig. 2, but both are made in a wide variety of patterns to meet many different circuit requirements. One type, the "twin" wire block, is made up of movable wires arranged in pairs, and may be used for either "make" or "break" contacts depending on the position in which it is assembled in the relay and the way the terminal ends are formed. These twin wire assemblies consist of a maximum of twelve pairs of round wires held in place in a molded phenolic plastic section near one end, and with contacts welded to the other. The assemblies differ in the number of contacts needed for the particular application of the relay, and in the position of the wire pairs across the width of the molded section. In the assembled relay, one of these twin wire blocks is mounted on each side of a stationary "single" wire block, the other basic type of assembly. This type consists of 12 relatively heavy wires, also molded permanently into plastic sections, but with different types and different numbers of contacts welded to the ends of the wires.

To obtain the wide range of combinations required, including make, break, sequence, and nonsequence contacts in various positions, it is necessary to have complete flexibility in machines and tools with arrangement for quick changeover from one contact pattern to another. Also, dimensional tolerances must be closely controlled to maintain the required closure intervals, operating speed, etc.

Wire Straightening

The first step in making these wire spring subassemblies is the accurate straightening of continuous lengths of wire. Special automatic multiple-head straightening machines have been developed at Western Electric for this purpose. These machines are used directly in conjunction with hydraulic molding presses, Fig. 3, where the plastic sections are molded around the straightened wires. The straightening machines operate at a speed slightly greater than the wire consumption of the molding dies, and automatic controls start and stop the straightener according to the demand at the press. Two sizes of nickel-silver hard-temper wire are used; 0.0226 inch in diameter for the twin wire blocks and 0.040 inch in diameter for the single-wire block. The straightening machines have been developed so that the design is essentially the same for wire of either size, differing only in the contour of the removable blades. This assures flexibility in setup as well as greatly simplifying maintenance and repair.

A separate spool of wire is provided for each wire required in a molded block. A capstan with individual pulley adjustment for each wire pulls the wire from the supply spools and guides it between drive rolls which push the wires through tubes to rotating straightening heads; see Fig. 4. The straightening heads are tubular and are separated in the middle so that each half can be rotated in opposite directions. Inside the head and rotating with it is a removable spindle with a free coupled joint in the middle; see Fig. 5. Held in each half of the spindle is a pair of precisely contoured die blades, which flex the wire and burnish the surface as they rotate around the wire. Halfway through the head the direction of rotation is reversed, so that the wire emerges without torsional stress, but with sufficient longitudinal surface stress to resist bending. The absence of torsional stress is of equal importance with straightness, the reason for which will be discussed later.

Automatic Molding

The wire blocks are molded in progressive dies, Fig. 6, four at a time, on two parallel arrays of either 12 or 24 wires. When making the 12-wire single block, half the wires from a 24-head straightener are guided into each forward cavity. When the 24-wire twin blocks are being made, the entire production of a 24-head straightener is guided into each forward cavity, and two machines are set up at each molding press.

The first station of the progressive molding die serrates or flattens a small section of the wires to provide a means of anchoring them in the plastic. Phenolic preforms, previously briquetted from powder in a mechanical tabletting press, are electronically preheated and automatically fed into the die beneath a transfer ram at the second station, where the actual molding is done. After the plastic is sufficiently cured under heat and pressure, the die opens and the cluster of four blocks is ejected from the mold and indexed to the third station where the blocks are clamped between water-cooled plates and the center cull sheared out. The fourth and final operation in the molding cycle is shearing the wires between the molded sections to form individual wire blocks.

During the entire series of molding operations, the wires are held taut continuously and kept perfectly aligned from the time they enter the guide slots until the completed blocks are finally cut apart at the end of the series. Ordinarily, no operator's attention to individual wires is necessary until one of the spools is empty.

Paper 56-650, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 23, 1956; made available for printing April 19, 1956.

JAMES W. RICE is with the Western Electric Company, Inc., Chicago, Ill.

Credit is due the many people who assisted individually in this development. Engineers at both the Western Electric Company and the Bell Telephone Laboratories co-operated fully and important contributions were made by both organizations.

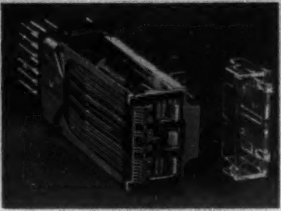


Fig. 1. **Wire spring relay, with contact cover removed**

At that time, the wire straightener is rolled away and the wires from a refilled straightener are threaded through the machine.

Press Controls

The electric controls for all the press operations are housed in a cabinet adjacent to the molding machine. The operation of the hydraulic press rams, the electronic preform heater, the feed mechanism, and the pneumatic devices on the die are all precisely timed and perfectly co-ordinated. Any operational sequence or length of cycle can be established for repetitive manufacture. On the other hand, turning a switch will place the entire operation on manual control for setup or maintenance.

Flash Removal

Even with the most careful die construction, the high molding pressure sometimes forces bits of plastic into the crack at the parting line of the die or out of the cavity along the wires. This leaves thin, brittle fins or "flash" which must be removed, lest particles break off later and lodge between the electric contacts. This flash is removed in special conveyorized grit-blasting machines; see Fig. 7. Here the blocks are loaded into fixtures, carefully designed to shield the fine molded wire guide teeth, but to expose the areas where flash is likely to occur. The blocks are then carried past nozzles where compressed air laden with ground walnut shells blasts the flash away. As the blocks are removed, they are placed in special slotted magazines, and not handled individually beyond this point.

Automatic Transfer Lines

The second major sequence of operations in the manufacture of these wire spring relay parts is perhaps more impressive than the wire-straightening, molding and flash removal just described. These are the remainder of the operations on the molded wire blocks which transform them into complete subassemblies, ready for final assembly into the relays.

Forming the contact end of the wires, cleaning, welding contacts in any pattern, coining the contacts to close tolerance, twisting and forming the terminals, clipping to exact length, tension-bending the wires, and removing unused wires are all done in integrated lines of machines through which the parts are transferred and processed completely automatically; see Fig. 8. As the parts come from the molding room they are stacked in a 6-position loading turret, at the left-hand end of the line. A shuttle-type transfer rail extending the length of the machine pulls out the bottom block onto the transfer track, leaves it in position and returns for the next part. Removal of the last block in the stack signals the turret to index to the next position.

At this time, perhaps some of the critical dimensions of the relay should be reviewed. These dimensions, as well as the way the parts fit together in the final assembly, govern the way the wire blocks are located at the work stations in the automatic line. In the assembled relay, the single-wire block is located between the two twin wire blocks, and pairs of precision dowels, molded as part of the plastic block, pass through mating holes in the twin wire blocks. This means of permanent and precise alignment is an important feature of the design for rapid assembly. All of the critical dimensions on the wire blocks are measured from the dowels or holes. At each work station in the transfer lines, the centers are first found and then the operations performed with constant dimensional relationship to them.

The first unit in the line twists the two wires of each pair together at the terminal side of the plastic block, which is an important operation in making the relay terminals suitable for solderless wrapped connections. This is actually a double unit and occupies two transfer track positions. Two identical heads with six spindles each twist alternate pairs of wires on two wire blocks with each operating cycle.

The second work station is a 35-ton press which coins the 12 twisted pairs into a trapezoidal cross section, see Fig. 9. This completes the solderless connection feature of the terminal and the two round wires are now joined into a rigid structure having the necessary sharp edges.

The third machine unit in the automatic line is a 8-ton hydraulic press equipped with a 2-stage progressive tool. Here is performed the most complex wire-forming operation in the entire line.

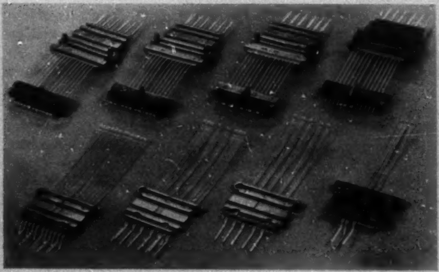


Fig. 2. **The two types of wire spring assemblies, showing various contact combinations**

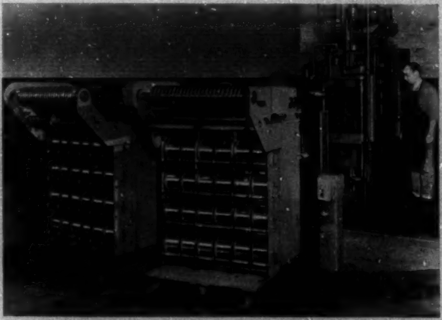


Fig. 3. **Wire straightening machines in conjunction with plastic molding press**

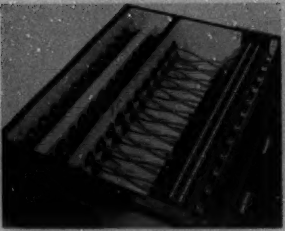


Fig. 4. Interior of 24-head wire straightening machine

Fig. 5. (right). Double spindle showing complement of two sets of removable die blades

The contact ends of the wires must be sharply offset to provide room for the actuating card immediately behind the contacts. The operation gives the wires a double reverse bend similar to the letter Z and at the same time the very tips of the wires are flattened and slightly curved on one side to form the base for the contact, and flattened to a plane on the opposite side to make a low resistance surface for the contact-welding operation. Also in this operation, the wires are flattened and slightly curved for a short distance immediately back of the offset bend where the actuating card will bear upon the wires. A sketch of this important wire shape is shown in Fig. 10. All of these dimensions must be held to close tolerances since they directly affect the contact-operate points. Here also is where the need for stress-free and nearly perfectly round wires becomes apparent. Release of stresses or rolling of the wires as the forming and flattening pressure is applied, will cause intolerable mislocation of the contacts. Fig. 11 shows the contact end of a relay having one twin wire block with torsionally stressed wires and the other without. The second station in the 2-stage tool clips the contact ends of the wires precisely and evenly.

A cleaning station is next in line, where a stream of liquid trichloroethylene washes all traces of dirt and oil off the wire ends where the precious-metal contacts are to be welded. At another station, also under the exhaust hood, a stream of compressed air dries the wires.

Twin Wire Contact Welding

The automatic twin wire contact-welding machine is the large unit near the middle of the line. Here a bit of previous metal is welded to the tip of each wire which, along with another precious-metal contact on the mating single wire, makes up the pair of surfaces which opens or closes the electric circuit when the relay operates. The contact material is palladium, 0.009 inch in thickness, with a 0.001-inch-thick layer of 22-carat gold fused to the top surface. This bimetal tape is about 0.030 inch in width and is fed to the machine in long lengths, wound on individual reels. On the bottom surface is a small bead, or ridge, which constitutes the "projection" for the welding process.

The wire blocks are transferred into the machine from the left side and precisely located with the reference holes in the plastic block engaged on pilot pins. The welding "head" itself consists of 12 electrodes, one for each alternate wire of the 24-wire block. An identical head is located farther along where contacts are welded on the even-numbered wires. As the block is positioned with the reference holes in the plastic engaged on pilot pins, a guide comb engages the individual wires near the plastic block, and then moves toward the ends of the wires, spacing and positioning them precisely along the top of the common lower electrode. Drive rolls feed the individual contact tapes forward the required distance (0.050 inch), and positions the ends over their respective wires. Cantilever suspended individual electrodes descend and press the tape ends firmly against the flattened surfaces of the wires, and the welds are made. The tapes are then clipped, and the part transferred to the second welding head where the process is repeated for the odd-numbered wires.

As was mentioned earlier, if the particular part being made will not require contacts in some positions, they will not be welded, and no precious metal will be wasted. The welding machine is completely versatile in this respect, since all 24 circuits are completely independent and the desired pattern is preset with the bank of toggle switches near the left side of the control panel, above the machine. The mechanical movements necessary to transfer and locate the part, operate the electrodes, feed, clamp, and clip the tape, all are controlled by a master cam shaft and thus kept in precise timing throughout the cycle.

The next operation on the twin wire block is the heaviest and also the most critical of the series. This is performed on a 15-ton press with a single station tool, so that the work station is directly under the ram, and distortion is minimized. This operation "builds in" the critical contact location dimension by coining the contact so that it has a smooth, curved surface, at the same time reflattening the card-bearing surface above the contacts.

The next operations are performed in an 8-ton press with a 2-station progressive tool. Here the terminal end of the blocks is completed by offsetting the wire pairs into staggered rows for easier wiring, and clipping the ends evenly. The last press is also 8-ton with a 2-station tool. The first station forms an offset bend in each wire so that the relay contacts will press against each other with the required force. The second station of this tool clips off any wires which are not to be used for the particular code of relay being run. This tool has a series of twelve clipping details which can be quickly arranged to clip or not clip in any pattern.

From this station the parts are carried to the end of the transfer track to the machine-unloading station where they are stacked in holding magazines. This station, like the loading station at the head end of the line, is also built on a turret but has eight stations instead of six. As each magazine is filled, the turret indexes, and brings an empty magazine into position. When six or seven magazines are

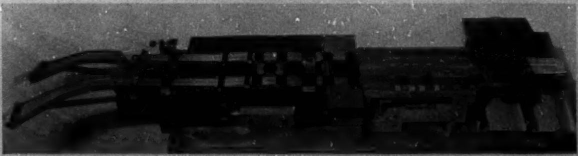


Fig. 6. Lower half of 4-cavity progressive molding die

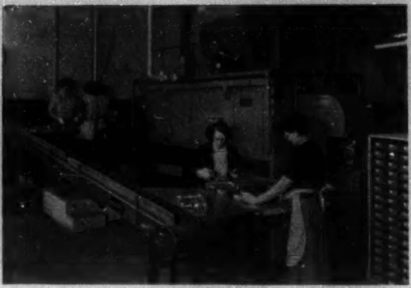


Fig. 7. Conveyorized grit-blasting machine for removal of molding flash

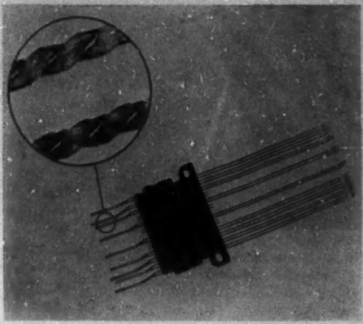


Fig. 9. Twin wire terminals twisted and coined for solderless wrapped connections

filled, the operator removes them and attaches empty magazines to the turret.

The control center is located above the welder, near the middle of the line. Here are located start and stop switches for each of the machine units as well as pilot alarms to show the location of any trouble condition. A relay cabinet is located farther down the line at the rear. The circuit is so designed that all the machines are electrically interlocked to keep them perfectly timed and to assure that all have completed their cycles before the succeeding cycle begins.

Tools and Presses

Several unique features have been incorporated into the designs of the press tools. With the reciprocating transfer method of feeding, the alignment of the transfer track with adjacent units is quite important. Each press tool is designed on the double-die set principle: a die set within a die set. The outer die set locates and supports the section of transfer track through the tool and is installed permanently in the press. The inner die set carries the intricate forming or cutting details and is mounted with precision screw dowels within the outer set. With this arrangement, the inner set can be quickly removed for maintenance, and replaced without disturbing the critical alignment, thus causing minimum down time for the line.

The double-die set principle and the general massiveness of the tool structure is also an important consideration in reducing distortion under pressing pressure. Any slight distortion is particularly objectionable when the operation is primarily one of squeezing to shape, as many of them are.

These operations of squeezing the wires to close tolerances also led to the selection of hydraulic presses rather than any type of crank press. Heavy stop blocks which locate the bottom of the stroke are designed into each press tool. The stops are ground to proper dimension in the tool room and any tool can be installed without adjustment of the press stroke.

Single-Wire Lines

Similar automatic transfer machinery and tooling for the final operations on the single-wire block has been developed and is in operation. Here also the blocks are handled in magazines from which they are mechanically unloaded, and transferred through the operations of precision clipping, contact welding, contact sizing, terminal forming, and tension bending, finally to be reloaded into magazines. The actual operations and tooling for the single blocks are, except for the contact welding, generally less complex than those for the twin wire blocks.

The only forming operations are those of serrating and offsetting the terminals and tension bending. Since the wires are heavier, individual rather than paired, and held in plastic at both ends, there is no problem of individually locating and guiding the wires at the work stations. The contact-welding machine, however, is altogether different from the twin wire contact welder.

Percussion Welding

The contacts on the single wire blocks are made up of small rectangular blocks of cupro-nickel with a 0.010-inch layer of palladium on one or both sides. These contact blocks are welded to the tips of the wires in various combinations; see Fig. 2. The design of the wire block is such that clamp electrodes, normally used in resistance welding and located near the weld zone, could not be used. A little-used welding technique, percussion welding, was developed for this application.

Simply stated, percussion welding involves the application of a high potential, 900 to 1,800 volts, across a gap between the parts as they move toward each other. The resulting arc heats the abutting surfaces before they meet, so that a thin layer of metal on each surface is brought to welding temperature, and

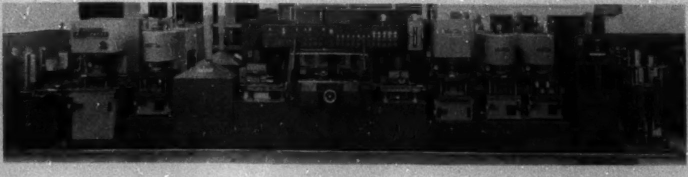


Fig. 8. Automatic transfer machine for contact-welding and wire-forming of twin wire blocks

Vol. 7
Pt. 1
1956

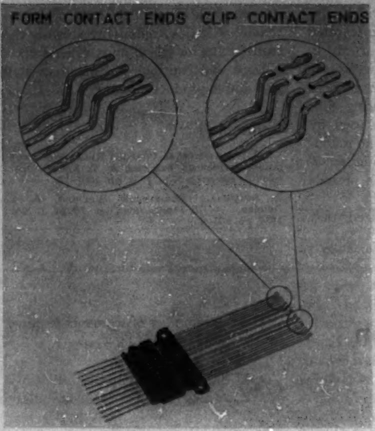


Fig. 10 (left). Sketch of formed and flattened end of twin wires before welding of contacts

Fig. 11 (above). Misalignment of relay contacts caused by torsional stresses or oval wires

the welding is made as the parts come together with an impact.

These are four contact conditions that may be applied to any of the 12 wires of the single-wire block; see Fig. 12. For a break relay contact, the palladium cap on the contact is on the bottom, and for a make contact, the cap is on the top. For transfer or sequence contacts, a palladium cap is required on both upper and lower sides. For relays requiring less than twelve contacts, some wires have the contact block omitted entirely. Three tapes, with palladium caps on the top side, the bottom side, and both sides, are fed from each side of the machine to two welding heads. The welder selects the type of contact needed, cuts it from the proper tape, and welds it to the proper wire. This information is preset by the setup man who has only to press one of four colored buttons associated with each of the 12 wire positions.

In operation, two wire blocks at a time are carried into position by the reciprocating transfer bar, located and clamped in a movable nest. During this process, the tape selector on each side of the machine has found the proper tape and has pushed it into a clipping die. Two punches move forward, clip contacts from the tapes, and carry them to transfer fingers which push them away from the punch and into movable jaws. The jaws are held on spring-propelled carriages, facing the first wire on each of the two nested wire blocks. One terminal of a charged capacitor is connected to the outside end of the wire through a movable brush and the other terminal to the jaw carriage through a flexible lead. At the proper instant, a cam releases the carriage which propels the jaw and contact toward the wire end. As the gap decreases, the arc establishes itself and heats the two abutting surfaces. As the surfaces meet, the arc is extinguished and the molten metal fuses. The jaw is pulled off the contact and returns to the loading position for the next contact while the wire blocks are indexed to the next wire position, where the process is repeated.

The wire blocks are handled two at a time and the two welding circuits are separate and independent. The capacitors are charged during the time the jaws are returning for their next contacts and the blocks are being indexed to the next wire position.

As stated earlier, and as shown in Fig. 2, some blocks do not require contacts in every position. The first percussion welders took care of this by not feeding the tape and cutting off the high voltage at these positions, but otherwise completing the entire mechanical cycle once for each of the 12 wires. A machine of a later model is now completed, which carries the blocks directly over the unused wire positions, stopping for the welding cycle only where a contact is required. This, of course, enables a marked increase in output, especially for those codes requiring only two or three contacts. This feature is completely automatic and any contact pattern may be obtained by depressing the proper buttons on the code selection panel.

The percussion welding machine of an earlier model has been described in a previous paper,[2] along with a discussion of many of the problems associated with the development of this unique welding process.

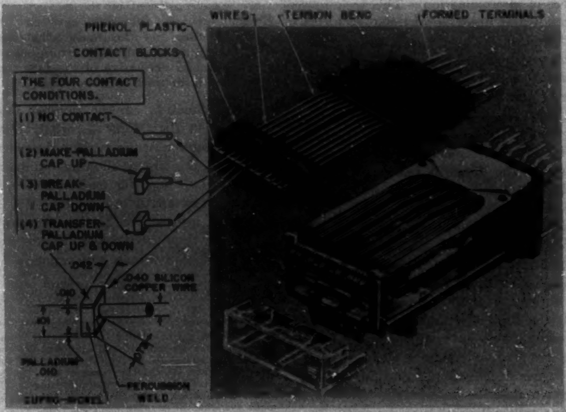


Fig. 12. Single-wire block with percussion-welded contacts. Also view of wire spring relay

Summary

This paper has described the manufacture of the major subassemblies for the general-purpose wire spring relay. Their mass production to close tolerances and at low cost has required the development of entirely new and different machines and techniques. Several machines of the type discussed are already in operation and more are under construction to satisfy an ever-increasing demand.

The combination of such widely different kinds of processes into integrated machine units is a great step forward in the march toward the automatic factory so widely acclaimed in these modern times. Knowledge gained during this development has already been used by Western Electric in planning the manufacture of future communication apparatus and in lowering the cost of present Bell System products.

It will doubtless be recognized that automatic facilities of the type described require a major capital investment, the justification for which must depend upon a large and continuing demand as well as a stable product design.

References

1. A NEW GENERAL-PURPOSE RELAY FOR TELEPHONE SWITCHING SYSTEMS, A. C. Keller. *Electrical Engineering*, Nov. 1952, pp. 1007–12.

2. AUTOMATIC PERCUSSION WELDING, A. L. Quinlan. *AIEE Transactions*, vol. 73, pt. I, Nov. 1954, pp. 561–65.

Optimum Linear-Segment Function Generation

HOWARD HAMER
ASSOCIATE MEMBER AIEE

DURING the past few years, the accuracy of linear analogue-computing components has been improved to the extent that the present limitation on solution accuracy in analogue computation and simulation is caused almost entirely by the errors in nonlinear equipment. One nonlinear device which has found extensive use is the diode-type generator for arbitrary functions. In general, this device approximates a given function by means of short linear segments, which are located in the function plane as closely as possible to the desired function to obtain a minimum error of approximation. This technique is known as linear approximation.

Fig. 1 shows a typical linear approximation to an arbitrary function. The point at which one straight-line segment meets the following segment is known as the break point for the succeeding segment. If both the slope and break point for each segment are continuously variable, the device is known as a variable-base function generator. As can be seen from Fig. 1, there are an infinite number of adjustments possible for such a generator, each of which approximates the function to some degree. Only one of these, however, is optimum in the sense that the maximum error between the given function and the approximation is as small as possible for the given function and a given number of segments. It is the purpose of this paper to describe an analytic technique which makes it possible to determine the smallest error obtainable and the break points and slopes which will achieve the optimum approximation.

Mathematical Development

In the Appendix it is shown that, if a section of a curve $y = f(x)$ is approximated by a straight-line segment (see Fig. 2)

$$(\Delta x)^2 = \frac{8E}{|y''|} \tag{1}$$

where

(Δx) = the length of the straight-line segment projected on the x-axis
E = the greatest difference between the function and the straight-line segment, and is called the error
y'' = the second derivative of the function, and is assumed constant over (Δx)

If the function can be divided into regions such that y'' is essentially constant throughout each region, then, for a given E, the largest allowable value of (Δx) can be found for each region. If a region is of length C_kX_m, where X_m is the total interval of the independent variable x, then the number of segments required to approximate this region is

$$N_k = \frac{C_kX_m}{(\Delta x)_k} \tag{2}$$

The total number of segments required for the entire curve is

$$N_t = \sum_{k=1}^{n} \frac{C_kX_m}{(\Delta x)_k} \tag{3}$$

If all the regions are of equal length, then C_k is a constant, equal to $1/n$, so that

$$N_t = \frac{X_m}{n}\sum_{k=1}^{n} \frac{1}{(\Delta x)_k} \tag{4}$$

If the end points and mid-point of a region are x_ky_k, $x_{k+1}y_{k+1}$, $x_{k+2}y_{k+2}$, then $x_{k+2} - x_k = X_m/n$ and

$$y'' \cong \frac{y_k + y_{k+2} - 2y_{k+1}}{(x_{k+1} - x_k)^2} \tag{5}$$

Since $x_{k+1} - x_k = {}^1/_2\,(x_{k+2} - x_k)$

$$y'' \cong \frac{4n^2(y_k + y_{k+2} - 2y_{k+1})}{X_m{}^2} \tag{6}$$

Let $y_{k+1} = y_k + \Delta_k{}^1$, $y_{k+2} = y_{k+1} + \Delta_{k+1}{}^1 = y_k + \Delta_k{}^1 + \Delta_{k+1}{}^1$. Then

$$y_k + y_{k+2} - 2y_{k+1} = \Delta_{k+1}{}^1 - \Delta_k{}^1 = \Delta_k{}^2 \tag{7}$$

where

$\Delta_k{}^1$ = the first difference of the k^th^ region
$\Delta_k{}^2$ = the second difference of the k^th^ region

In general

$$y_k + y_{k+2} - 2y_{k+1} \triangleq \Delta_k{}^2 \tag{8}$$

Thus

$$y'' \cong \frac{4n^2\Delta_k{}^2}{X_m{}^2} \tag{9}$$

or

$$(\Delta x)_k{}^2 = \frac{8EX_m{}^2}{4_n{}^2|\Delta_k{}^2|} \tag{10}$$

$$\frac{1}{(\Delta x)_k{}^2} = \frac{n^2|\Delta_k{}^2|}{2EX_m{}^2} \tag{11}$$

$$\frac{1}{(\Delta x)_k} = \frac{2n\sqrt{|\Delta_k{}^2|}}{\sqrt{8EX_m}} \tag{12}$$

also

Paper 56-696, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted May 17, 1955; made available for printing April 26, 1956.

HOWARD HAMER is with Electronic Associates, Long Branch, N. J.

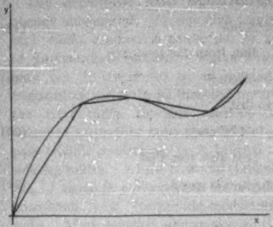


Fig. 1. A typical linear approximation

$$(\Delta x)_k^2 = \frac{2E(x_{k+2}-x_k)^2}{y_k+y_{k+2}-2y_{k+1}} \tag{12A}$$

Substituting equations 12 and 12(A) in equation 4

$$N_t = \frac{X_m}{n}\sum_{k=1}^{n}\frac{2n\sqrt{|\Delta_k^2|}}{\sqrt{8E}\,X_m}$$

$$= \frac{1}{\sqrt{2E}}\sum_{k=1}^{n}\sqrt{|\Delta_k^2|} \tag{13}$$

Solving for EN^2

$$EN^2 = \frac{1}{2}\left(\sum_{k=1}^{n}\sqrt{|\Delta_k^2|}\right)^2 \tag{14}$$

which is the equation relating error and number of segments for the optimum linear approximation to an arbitrary function.

Applications

Equation 14 can be used to find the approximation error if the number of segments is given, or to find the number of segments required for a desired accuracy. The total range of the independent variable x is divided into n intervals, where n is arbitrary, but should be as large as calculation time permits. The function is tabulated at the end points of these intervals and also at the mid-points. Table I is a tabulation for $y = \epsilon^{-x}$ for which 10 intervals are used. The second differences are computed for these 10 points, the square roots taken and summed. When this number is put into equation 14

$$EN^2 = 1/2(0.3934)^2 = 0.0774$$

If it is desired to approximate this function with five segments

$$E = 1/25 \times 0.0774 = 0.0031 \text{ or } 0.31 \text{ per cent}$$

If, however, an accuracy of only 1.0 per cent is required, with the use of a minimum number of segments, e.g., a special loaded potentiometer, perhaps, where segments cost money

$$N = \sqrt{\frac{0.0774}{0.01}} = 2.8$$

This analysis makes no provision for the technique known as error-splitting, whereby the approximating linear segment is moved vertically by half the error and results in a final error which is half, approximately, of that calculated by equation 14. If this technique is to be used, and it is almost always advisable, the value for E in equation 3 must be twice the final desired or expected error.

Once equation 14 has been used, and N and E are determined, equation 12(A) can be used to locate the required end points for the optimum approximation. A point x_1 is chosen which is about where the first segment is expected to end. This defines a region

$$x_0 \leq x \leq x_1$$

Equation 12(A) is applied to this region, yielding $(\Delta x)_1$, which is the length of the first interval. A second region

$$x_0 + u_1 \leq x \leq x_2$$

is then chosen, equation 12(A) applies, and $(\Delta x)_2$ is found, which defines interval 2 as

$$x_0 + u_1 \leq x \leq x_0 + u_1 + u_2$$

This process is repeated until the range of the independent variable is exhausted. The total number of segments required should be the same as that used in, or obtained from, equation 14. Once the segment lengths have been determined it is a simple matter to set up the approximating function.

Example: $y = \epsilon^{-x}$ in 5 segments.

Table I. Tabulation for $y = \epsilon^{-x}$

n	x	y	Δ^1	$\lvert\Delta^2\rvert$	$\sqrt{\lvert\Delta^2\rvert}$
1.	0.0	1.00000	0.04877	0.00238	0.0488
	0.05	0.95123	0.04639		
2.	0.10	0.90484	0.04413	0.00215	0.0464
	0.15	0.86071	0.04198		
3.	0.20	0.81873	0.03993	0.00195	0.0442
	0.25	0.77880	0.03798		
4.	0.30	0.74082	0.03613	0.00176	0.0420
	0.35	0.70469	0.03437		
5.	0.40	0.67032	0.03269	0.00159	0.0399
	0.45	0.63763	0.03110		
6.	0.50	0.60653	0.02958	0.00144	0.0379
	0.55	0.57695	0.02814		
7.	0.60	0.54881	0.02676	0.00130	0.0361
	0.65	0.52205	0.02546		
8.	0.70	0.49659	0.02422	0.00118	0.0344
	0.75	0.47237	0.02304		
9.	0.80	0.44933	0.02191	0.00106	0.0326
	0.85	0.42742	0.02085		
10.	0.9	0.40657	0.01983	0.00097	0.0311
	0.95	0.38674	0.01886		
	1.00	0.36788			

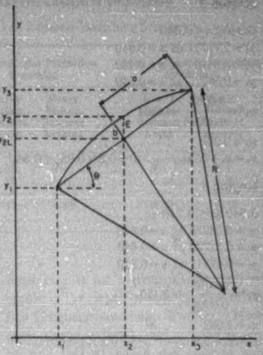


Fig. 2. Linear approximation to a short segment of curve

From Table I

$$\sum_{k=1}^{10}\sqrt{|\Delta_k^2|} = 0.3934$$

Step 1

From equation 14

$$EN^2 = \frac{(0.3934)^2}{2} = 0.0774$$

If $N = 5$

$$E = \frac{0.0774}{25} = 0.0031$$

With error-splitting

$E_s = 0.00155$ or 0.155 per cent

Step 2

Estimate region 1 from $x = 0$ to $x = 0.2$

$$u_1^2 \triangleq \frac{(\Delta x)^2}{(x_3-x_1)^2} = \frac{2E}{y_1+y_3-2y_2}$$

from equation 12A.

$y_1 = 1.00000$ at $x = 0$
$y_2 = 0.90484$ at $x = 0.1$
$y_3 = 0.81873$ at $x = 0.2$

$$u_1^2 = \frac{0.0062}{0.00905} = 0.685$$

$u_1 = \sqrt{0.685} = 0.828$
$x_1 = 0.828 \times 0.2 = 0.166$
Use $x_1 = 0.160$.

Region 2

$0.160 \leq x \leq 0.360$
$y_1 = 0.85214$ at $x = 0.16$
$y_2 = 0.77105$ at $x = 0.26$
$y_3 = 0.69768$ at $x = 0.36$

$$u_2^2 = \frac{0.0062}{0.00772} = 0.803$$

$u_2 = 0.896$
$x_2 = x_1 + 0.2u_{n1} = 0.16 + 0.2 \times 0.896 \cong 0.340$

Region 3

$0.340 \leq x \leq 0.540$
$y_1 = 0.71777$ at $x = 0.340$
$y_2 = 0.64404$ at $x = 0.440$
$y_3 = 0.58275$ at $x = 0.540$

$$u_3^2 = \frac{0.0062}{0.00644} = 0.964$$

$u_3 = 0.981$
$x_3 = x_2 + 0.2u_{n3} = 0.340 + 0.2 \times 0.981 \cong 0.540$

Region 4

$0.540 \leq x \leq 0.740$
$y_1 = 0.58275$ at $x = 0.540$
$y_2 = 0.52729$ at $x = 0.640$
$y_3 = 0.47711$ at $x = 0.740$

$$u_4^2 = \frac{0.0062}{0.00528} = 1.175$$

$u_4 = 1.082$
$x_4 = x_3 + 0.2u_{n4} = 0.540 + 0.2 \times 1.082 \cong 0.770$

Region 5

$0.770 \leq x \leq 0.970$
$y_1 = 0.46301$ at $x = 0.770$
$y_2 = 0.41895$ at $x = 0.870$
$y_3 = 0.37908$ at $x = 0.970$

$$u_5^2 = \frac{0.0062}{0.00419} = 1.48$$

$u_5 = 1.22$
$x_5 = x_4 + 0.2u_{n5} = 0.770 + 0.2 \times 1.22 \cong 1.000$
$y_5 = 0.36788$

The resultant end points which will approximate $y = \epsilon^{-x}$ to 0.15 per cent from $x = 0$ to $x = 1$ are:

x	y
0	0.99845
0.160	0.85059
0.340	0.71022
0.540	0.58120
0.770	0.46146
1.000	0.36633

Appendix

In Fig. 2 a short segment of a curve is shown, together with a straight-line approximation to the curve, intersecting at end points x_1, y_1 and x_3, y_3. It is assumed that the average radius of curvature of the segment of curve is R so that the curve is very similar to an arc of a circle of radius R passing through the end points and an intermediate point x_2, y_2. The corresponding intermediate point on the linear segment is x_2, y_{2L}, and the error of approximation is $E = y_2 - y_{2L}$. The angle of slope of the linear segment is θ and, if the preceding assumptions hold

$$\tan\theta \cong \left(\overline{\frac{dy}{dx}}\right) \tag{15}$$

where $(\overline{dy}/dx)$ is the average first derivative of the function over the interval. From the geometry

$$a^2 + (R-b)^2 = R^2 \tag{16}$$

or

$$a^2 + R^2 - 2Rb \cong R^2 \tag{17}$$

so that

$$a^2 \cong 2Rb \tag{18}$$

But, from the geometry

$$b = E\cos\theta \tag{19}$$

so that

$$a^2 \cong 2RE\cos\theta \tag{20}$$

It is also seen that

$$x_3 - x_1 = 2a\cos\theta \tag{21}$$

or

$$a \cong \frac{x_3 - x_1}{2\cos\theta} \cong \frac{\Delta x}{2\cos\theta} \tag{22}$$

where $\Delta x \triangleq x_3 - x_1$.
The definition of radius of curvature is

$$R \triangleq \frac{[1+(y')^2]^{3/2}}{|y''|} \tag{23}$$

where y' is the first derivative of $y = y(x)$, and y'' is the second derivative.

Since y' also equals $\tan\theta$

$$[1+(y')^2]^{1/2} = \frac{1}{\cos\theta} \tag{24}$$

$$R = \frac{1}{|y''|\cos^3\theta} \tag{25}$$

Substituting equations 22 and 25 into equation 20

$$\frac{(\Delta x)^2}{4\cos^2\theta} = \frac{2E\cos\theta}{|y''|\cos^3\theta} \tag{26}$$

$$(\Delta x)^2 = \frac{8E}{|y''|} \tag{27}$$

Automatic Protection Switching for TD-2 Radio System

G. H. KLEMM
ASSOCIATE MEMBER AIEE

Synopsis: Intercity transmission of television programs and telephone conversations is now provided via a network of *TD-2* microwave radio relay equipment operating on a transcontinental basis. Whenever service interruption threatens because of equipment troubles or atmospheric disturbances in the transmission path, a switching system automatically and quickly substitutes a stand-by wide-band microwave radio channel for the defective channel. The circuits and operation of this switching equipment are described.

TD-2 Microwave Radio Relay System

The *TD-2* radio system is designed especially for long-haul backbone-route service between major population centers, but it is also adaptable to light-route applications. Radio repeater stations are frequently located on high land and generally use towers or other high antenna supports to provide clear line-of-sight transmission paths which average approximately 30 miles between stations, depending upon the intervening terrain. The information is transmitted via very sharply focused microwave radio beams directed from one repeater station to the next.

Six 2-way broad-band radio channels are provided on a fully loaded *TD-2* system and any one channel will carry several hundred telephone circuits or one television signal. Each direction of transmission is independent of the other. The six channels for one direction of transmission use common transmitting and receiving antennas, and each channel uses independent electronic equipment. Each repeater station contains a common power plant. Batteries floating across the output of this power plant provide service in the event of short-time primary-source power failures and, in the event of long-time primary-source power failures, an engine-generator set is switched into service. The fusing and discharge leads are arranged so that single equipment failure will affect no more than one radio channel in one direction of transmission. Repeater stations are normally unattended but automatic alarms and remote controls are provided via wire lines to control centers. The latter are usually located at major population centers along the route or at other Bell Telephone System locations which are continuously manned.

Paper 56-698, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 28, 1956; made available for printing April 27, 1956.

G. H. KLEMM is with Bell Telephone Laboratories, Inc., Murray Hill, N. J.

The microwave radio signal is frequency modulated. At repeater points, each radio channel is converted to an i-f (intermediate frequency) signal which may be: 1. connected to an outgoing transmitter and antenna directed toward the next radio repeater station; 2. switched to another radio channel to bypass trouble conditions; or 3. bridged to side-leg routes. At the terminal stations, the i-f signal is converted to its original base-band a-m (amplitude-modulated) form. More detailed information on the *TD-2* system may be found elsewhere.[1,2,3]

Table I. Percentage of Time that Regular Channel is Unavailable

4,000 Miles, Five Regular Channels, Ten Repeater Switching Sections

	No Automatic Protection Switching		With Automatic Protection Switching	
	Heavy Fading, Summer Night	Heavy Fading, Summer Evening	Heavy Fading, Summer Night	Heavy Fading, Summer Evening
Equipment failure	6.65	6.65	0.1	0.1
Fading	1.46	0.03	0.13	0.02
Total	8.11	6.68	0.23	0.12
Automatic switching improvement factor			35	56

Service Failures

Service via radio equipment may be interrupted by equipment failures or by radio fades. Some equipment troubles are relatively slow (1 or more seconds) in causing serious degradation in service, e.g., when an electron tube heater fails, while some equipment troubles may interrupt service instantaneously, as in the case of the breakdown of a capacitor or opening of a defective connection. The use of high-quality components and preventative maintenance on a routine basis reduces equipment failure to a minimum.

The atmosphere through which the sharply focused microwave radio beam propagates from one repeater to the next is a dielectric with continuously varying characteristics. The beam is often divided and travels by more than one path so that the signal seen by the receiver is a composite of many signals all of different amplitude and phase. The magnitude of the received signal is continuously varying and this condition is referred to as fading. Under extreme conditions, the fade can cause complete service failure.

The effect of a fade in the *TD-2* radio system is an increase in noise in the telephone and television circuits. A deep fade may cause the carrier-to-noise ratio at the input to a repeater to approach unity. As the fade grows still deeper, the signal-to-noise ratio will drop off still more rapidly for a given db (decibel) increase in the depth of fade because of the f-m (frequency-modulated) effect. However, this so-called breaking point is well beyond that at which the noise is considered to be intolerable. Radio systems are designed to provide transmission margin sufficient to protect against fades of moderate depth. In those cases where the fades are deep enough to cause service failure, transmission usually returns to normal in a short time.

Available studies[4] indicate that the probable duration of interruptions of service caused by equipment failures on a 4,000-mile transcontinental 1-way channel of the *TD-2* radio system is 6.65 per cent of the time or 95 minutes per day.

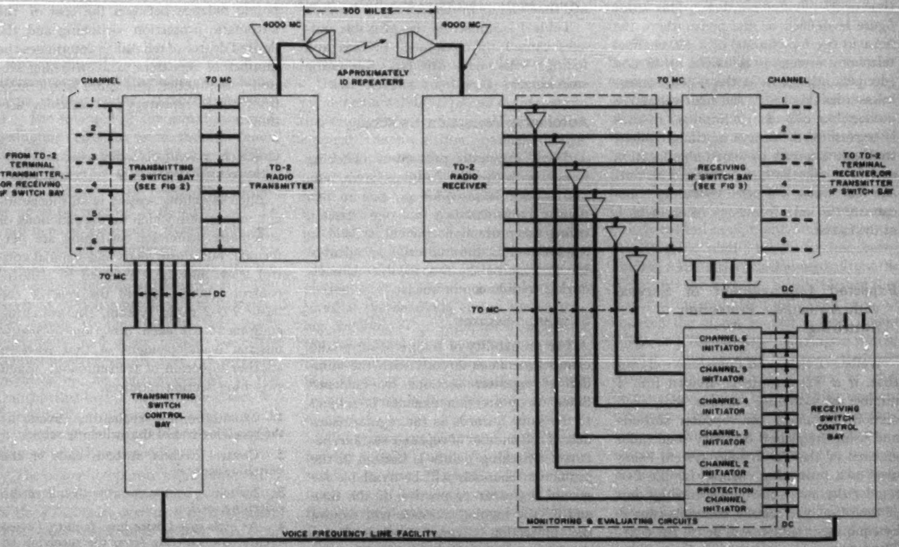

300 MILES
4000 MC
4000 MC
APPROXIMATELY 10 REPEATERS
CHANNEL
70 MC
1
2
3
4
5
6
FROM TD-2 TERMINAL TRANSMITTER, OR RECEIVING IF SWITCH BAY
TRANSMITTING IF SWITCH BAY (SEE FIG 2)
TD-2 RADIO TRANSMITTER
TD-2 RADIO RECEIVER
70 MC
CHANNEL
RECEIVING IF SWITCH BAY (SEE FIG 3)
TO TD-2 TERMINAL RECEIVER, OR TRANSMITTER IF SWITCH BAY
70 MC
DC
TRANSMITTING SWITCH CONTROL BAY
DC
TO MC
CHANNEL 6 INITIATOR
CHANNEL 5 INITIATOR
CHANNEL 4 INITIATOR
CHANNEL 3 INITIATOR
CHANNEL 2 INITIATOR
PROTECTION CHANNEL INITIATOR
RECEIVING SWITCH CONTROL BAY
MONITORING & EVALUATING CIRCUITS
DC
VOICE FREQUENCY LINE FACILITY

Fig. 1. Automatic switching

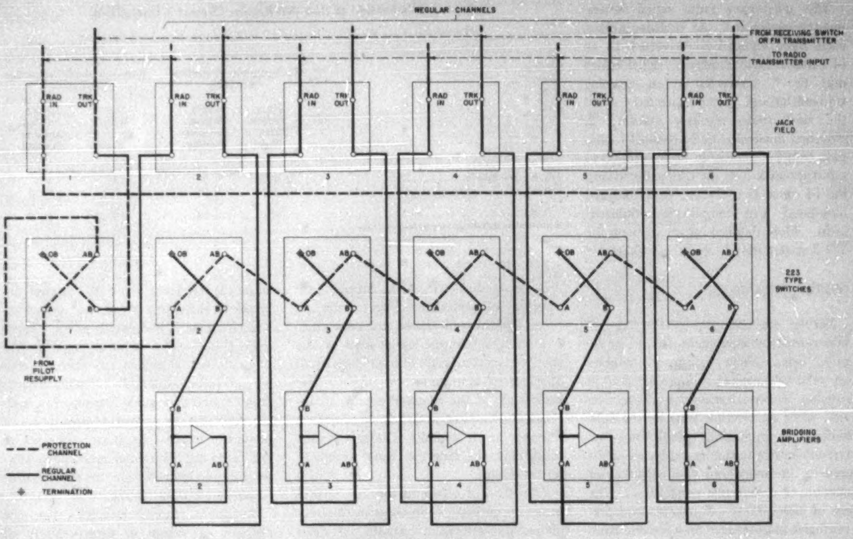


Fig. 2. TD-2 auto-switching transmitting switch

Vol. 7
Pt. 1
1956

A similar figure caused by fading in the worst fading period is 1.46 per cent of the time or 21 minutes per day. This latter figure is defined as the point where the noise in the top channel of a 480-channel telephone system is in excess of 40 dba (decibels adjusted) at the −9-db transmission level point. The relatively large outage time caused by equipment failures is because of the length of time required to dispatch a man to the trouble location and to repair the equipment. The outage time which is caused by fading is because of the large frequency of occurrence of the fades.

Expected Improvement of Service with Automatic Protection Switching

Welber, Evans, and Pullis[4] have shown that, if a *TD-2* route is divided into a number of switching sections, with each section containing ten repeater stations and using frequency diversity (one radio channel of the *TD-2* radio system being used as a protection channel for the five regular channels), the outage time for any of the other five regular channels caused by equipment failures will be 0.1 per cent. The outage time caused by fading will be 0.13 per cent during the summer night's heavy fading conditions. These figures include provision for the probable outage time for the protection radio channel.

Table I summarizes the probable outages caused by equipment failures and fading with and without automatic switching.

Automatic Protection Switching

When automatic protection switching is used with the *TD-2* radio system, normal service is provided on one to five regular radio channels and one channel, called the protection channel, is held in reserve and is automatically substituted for any one of the five regular channels during trouble conditions.

SWITCHING SECTION

The probability of failure of any radio channel increases directly with the number of repeater sections in tandem.[4] Since the protection channel is subject to the same hazards as the regular channels, if the number of repeater stations between switching points is limited, 1. the protection channels will be available for service a greater proportion of the time and 2. the regular channels will require the protection channel services for a smaller proportion of the time. Thus, by dividing the *TD-2* route into a number of switching sections, the reliability of a long route will be increased. An economic balance between the cost of the automatic protection switching and the desired degree of reliability determines the number of repeaters in a switching section. In practice, this figure is in a range from 5 to 15 repeaters per switching section.

CIRCUITS AND EQUIPMENT OF A SWITCHING SECTION

Each switching section is an independent unit within which all the functions of automatic protection switching are performed, but alarm leads and manual control leads may be connected to remote-control centers. As in the case of the basic *TD-2* radio system, the switching equipment for each direction of transmission is independent of that used in another direction of transmission. Each switching section requires:

1. Monitoring and evaluating circuits at the receiving end of the switching section.
2. Control circuits at both ends of the switching section.
3. Switching equipment at both ends of the switching section.
4. A high-speed wire-line facility (voice circuit) transmitting from the receiving to the transmitting end of the switching section.

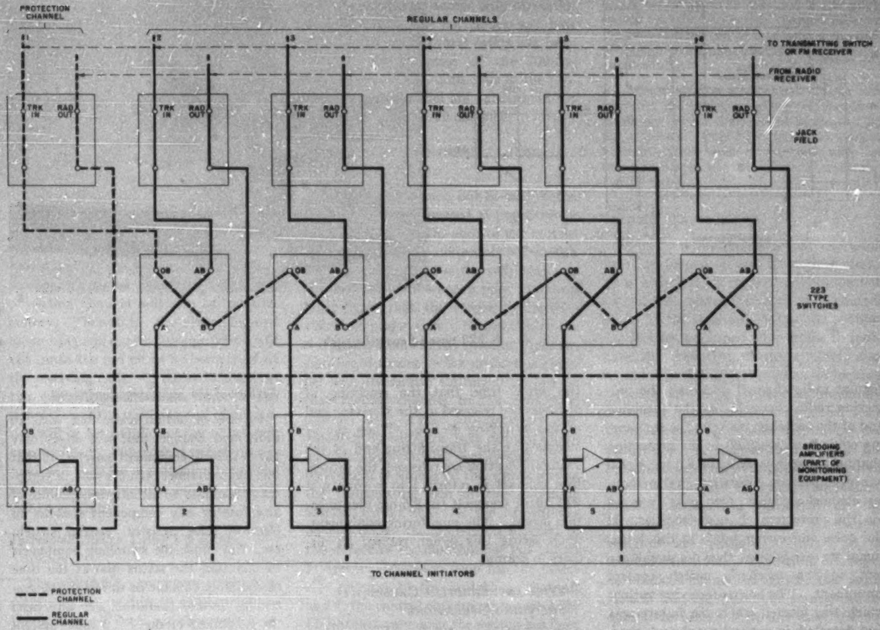


Fig. 3. TD-2 auto-switching receiving switch

Fig. 1 shows how the equipment is interconnected. All switching and monitoring is done at the 70-mc (megacycle) i-f. The function of the monitoring and evaluating circuits is to monitor each radio channel at the receiving end of the switching section and to call for the protection channel to be bridged into service or to be released from service, as the conditions of the regular channels change. D-c signals which indicate the condition of each radio channel are passed to the receiving switch control equipment. The receiving switch control equipment passes d-c signals to the receiving i-f switch equipment and also passes voice-frequency control signals over the wire-line facilities to the control circuits at the transmitting end of the switching section. The latter control circuits pass d-c signals to the local i-f switches in the transmitting switching equipment.

The power to operate this equipment is obtained from the common power plant of the *TD-2* radio system at each location. The same safeguards which relate to power or equipment failures are also employed in this case.

Operation of a Switching Section

Under normal conditions, one to five regular channels of the *TD-2* radio system operate in their normal manner and one channel, the protection channel, carries an 8.5-mc identification signal. This pilot signal is above the normal base-band information frequencies transmitted by the *TD-2* radio system and is inserted into the protection channel at the transmitting end of each switching section.

Each radio channel, including the protection channel, is continuously monitored at the receiving end of the switching section. This equipment monitors the noise level in a narrow band above the base-band information on each channel. Trouble in any channel is indicated by an increase in this noise level or a loss of carrier.

The protection channel is substituted for any regular channel by the following series of events. The monitoring equipment sees an increase in noise level above the threshold value required for a switch. D-c signals are sent to that particular area of the receiving switch control equipment associated with the regular radio channel in trouble. The receiving switch control equipment prepares to operate the receiving coaxial i-f switches which will substitute the output of the protection channel for the output of the failed regular channel. However, the output of the protection channel is still without base-band information except for the 8.5-mc identification signal. The receiving switch control equipment, which prior to the "trouble alert" condition has been sending a 700-cps (cycle-per-second) guard signal over the wire-line facility toward the transmitting end of the switching section, changes the signal to 900, 1,100, 1,300, 1,500, or 1,700 cps depending upon which regular radio channel has requested assistance. The control circuits at the transmitting end of the switching section recognize the change in the voice-frequency signal, know which regular channel is to be switched, open the normal connection on the protection channel, and bridge the input of the protection channel to the regular radio channel indicated by the voice-frequency wire-line signal. Thus, the same base-band information is now being sent over the

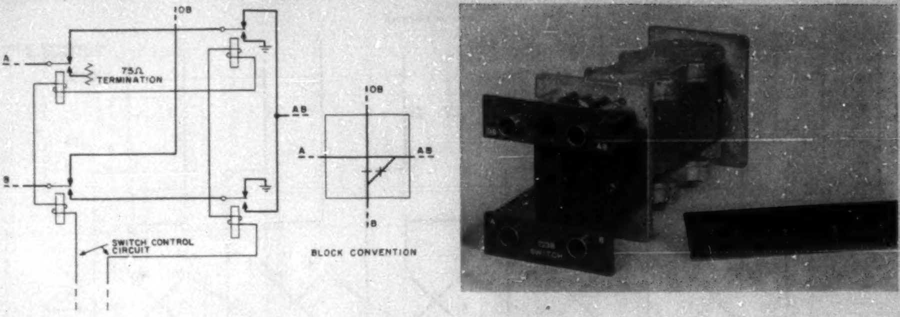


Fig. 4. A 223-type i-f coaxial switch

regular radio channel and over the protection radio channel. At the receiving end of the switching section, the monitoring equipment assigned to the protection channel recognizes that the 8.5-mc signal has disappeared, indicating that a regular channel signal is now being received via the protection channel facilities. If the noise and carrier levels in this latter signal are satisfactory, then d-c signals are sent to the receiving switch control equipment. The proper receiving switch, which was alerted when the failure was first detected, is now operated and, thus, the switch to the protection channel is completed.

The switching always takes place in the 70-mc coaxial circuits. In addition to connecting the desired signals to through paths, the otherwise open-ended coaxial circuits are terminated in the 75-ohm characteristic impedance of the coaxial equipment. During the progress of the switching operations described in the foregoing, alarms are sent to control centers and to the receiving end of the switching section.

The regular channel is returned to normal operation by the following series of events: When the regular channel becomes normal, the noise and carrier level return to normal and the d-c switch control instructions to the receiving control equipment are cancelled. The i-f switches at the receiving end of the switching section are released and the regular channel is returned to normal service. However, the protection channel is still bridged to the regular channel at the transmitting end of the switching section. This bridge is released by the transmitting switch control circuits when it receives a 700-cps guard tone over the wire line. This guard tone replaces the higher frequency voice-frequency signal on the wire line at the same time that the receiving i-f switches are released at the receiving end of the switching section. When the i-f switches at the transmitting end of the switching section are released, the protection channel is returned to normal and will again transmit its 8.5-mc identification signal. The monitoring equipment, upon seeing this signal, returns its circuits to normal.

SERVICE INTERRUPTION CAUSED BY SWITCHING OPERATIONS

The time required for a switch to complete is dependent upon the sum of the recognition and operate times of the receiving-end equipment, the transmitting-end equipment and the time of transmission of orders over the voice-frequency wire-line facilities. The maximum time objective for a switch to complete is 50 milliseconds and the usual switch is completed in 30 to 50 milliseconds. In the case of a sudden failure, the radio channel is open for this period of time. In the case of slower failures because of fades, etc., the switch request is started before the signal becomes unusable. Usually, the service will not be affected. During the more rapid fades, a television viewer may see the background noise of his picture increase and then suddenly disappear as the switch is completed.

SWITCHING SAFEGUARDS

The chaos which would be possible without safeguards to co-ordinate each operation of the switching equipment to allow for failure of the switching equipment or to prevent a single trouble from tying up a transcontinental route is obvious. The following illustrate the possible situations which have been allowed for in the design of the switching equipment.

Failure of the Switching Equipment

Failure of a fuse associated with an individual channel unit will affect only the switching equipment assigned to that one radio channel. In the case of equipment common to all channels, failure of any fuse or any component such as an electron tube, resistor, coaxial switch, etc., will cause the switching equipment to maintain the status quo at the time of equipment failure or to return the system to normal operation. In any case, the reliability of the *TD-2* radio system will not be less than it would have been without the automatic-protection switching equipment.

Simultaneous Failure on Two Channels

The switching equipment has only one protection channel available and cannot take care of two simultaneous troubles within one switching section. However, the probability of such troubles is reduced by limiting the number of repeater stations within a switching section. When two troubles do occur within a switching section, the first channel to get its request to the control circuits at the receiving end of the switching section takes priority. Once the protection channel is substituted for the first channel to request a switch, then no other switch orders will be accepted until the first regular channel has returned to normal service.

Single Failure Causing Switch in Subsequent Switching Sections

The switching equipment does not interchange information between switching sections. However, failure of the signal in any radio channel is seen by all subsequent switching sections. Every switching section will initiate a switch

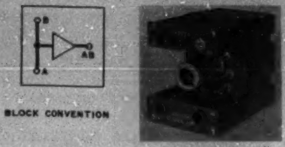


Fig. 5. I-f bridging amplifier

and, when the protection channel is bridged to the regular channel at the transmitting end of all switching sections subsequent to the section containing the trouble, the signal transmitted over the protection channel will also be unsatisfactory. Therefore, the receiving-end monitoring and evaluating equipment will not cause the switch to be completed at the receiving end. In the switching section containing the trouble, however, the signal seen at the receiving end of the section will be acceptable and the switch will be completed. Subsequent switching sections will then see a good signal over the regular channel and the switch requests will be cancelled.

Incomplete Switch Tying Up Entire System

The switching equipment must complete a switch within 50 milliseconds or the system will lock out that switch request for 10 seconds and accept any other switch request. This feature is known as recycling, which means that, if two switch requests are received and, for some reason, the first switch request cannot be completed, then, after 50 milliseconds, the first request will be ignored and the second request will be accepted as though no other request were present. In the event that no other second request is present at the end of the 10-second period, then the first request for a switch will be processed for 50 milliseconds. This routine is continued until either the first request is cancelled or the switch is completed.

CO-ORDINATION OF SWITCHES AT TRANSMITTING AND RECEIVING ENDS OF SWITCHING SECTION

The equipment at the transmitting end of a switching section is completely slave to the equipment at the receiving end of the switching section where all decisions are made. The receiving-end control equipment can only send one voice-frequency signal toward the transmitting-end equipment at a time, and the 700-cps guard tone is always transmitted before another identification voice-frequency signal is transmitted. The transmitting-end control circuits can accept only one order at a time and will not honor any new orders until the previous order has been cancelled by reception of the 700-cps guard tone. Thus, both ends of the protection radio channel are substituted for the same regular radio channel.

SWITCHING AT INTERMEDIATE POINTS ALONG THE ROUTE

In television service, one or more radio channels may be dropped at points along the route between the ends of the switching section. If a failure occurs between the transmitting end of the switching section and the point where a television drop takes place, then the service to the television drop also fails. No matter where the failure occurs in the switching section, transfer of service to the protection channel will necessitate shifting all drops on that regular channel to the protection channel also. These situations are provided for by what has been termed auxiliary station switching equipment. This equipment is similar to the equipment at the transmitting end of the switching section, except that: 1. the i-f switches do not bridge the protection channel to the defective regular channel but rather bridge the television drop for the defective channel to the protection channel; and 2. the control equipment actuates the i-f switches only after the switch has been completed on the over-all switching section and a supplementary 2,100-cps voice-frequency signal has been received over the wire line. The system is restored to normal by the same 700-cps guard tone used to restore normal conditions at the transmitting end of the switching sections.

ALARMS

Each trouble condition causes an alarm. These alarms are indicated at the receiving end of a switching section and may be duplicated at remote locations. A summary of the alarms are as follows:

1. A major alarm is operated when a regular channel fails and is not replaced by the protection channel within a few seconds.

2. Minor audible and visual alarms are operated to indicate failures on the regular or protection channels when the failure, lasts for more than 15 to 45 seconds. In the case of a regular channel failure, service must be provided via the protection channel. In the case of protection channel failure, all the regular channels must be in normal operating condition. The visual alarm indicates the number and direction of the channels in trouble.

3. A major alarm is also operated within a few seconds if the protection channel fails after it has been switched for an impaired regular channel or for use as an additional channel.

4. A major alarm is operated if the pilot supply for the protection channel fails at the transmitting end of a switching channel or if the pilot supplies used in the evaluating circuits fail.

5. Minor audible and visual alarms are operated to indicate failure of a voice-frequency path. Two alarm lamps are lit for a failure. One lamp indicates that a voice-frequency path is in trouble and the other indicates the direction of transmission affected.

MANUAL CONTROLS

Even automatic switching systems must be turned on and off and be manually switched for repairs and equipment or circuit rearrangements. A control panel at the receiving end of the switching section, equipped with lamps to indicate the condition of the circuits, also contains push-button switches for locking out the switching system or manually switching the system. This panel may be duplicated at remote points by d-c connection. A master button and the button representing the operation desired must be depressed together; this feature prevents accidental operation of the equipment. Various types of manual switches are possible. One type of manual switch requests a switch of a regular radio channel in much the same fashion that the automatic equipment requests the switch. In this case, the 50-millisecond timer is disabled and the switch will not be completed if service via the protection channel is not satisfactory. However, there are situations where, for service or other reasons, it is desired to switch definitely over to the protection channel. In this case, there is a forced manual switch button. Another type of switch involves the regular channel lockout which is requested when operating conditions warrant removing any regular channel from the protection of automatic switching. In other instances, traffic may be such that the protection channel is required for regular service in addition to the other five channels. In this case, the protection channel lockout button is depressed and, if the protection channel is not at that time being used as a substitute for a regular channel, then the automatic switching equipment is locked out of service and the reliability of the *TD-2* radio system reverts to that of the unprotected system.

EQUIPMENT AND CIRCUIT ARRANGEMENTS

The same basic building blocks are used at the transmitting end, at the receiving end, and at the auxiliary stations of a switching section. These include the

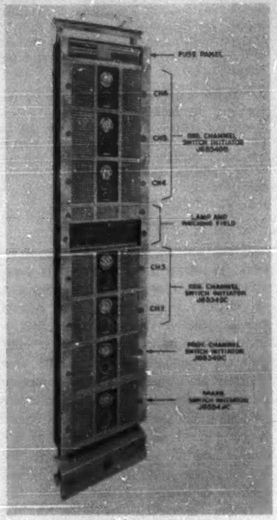

FUSE PANEL
CH6
CH5
REG. CHANNEL SWITCH INITIATOR J68349B
CH4
LAMP AND PATCHING FIELD
CH3
REG. CHANNEL SWITCH INITIATOR J68349C
CH2
PROT. CHANNEL SWITCH INITIATOR J68349C
SPARE SWITCH INITIATOR

Fig. 6. Initiator bay

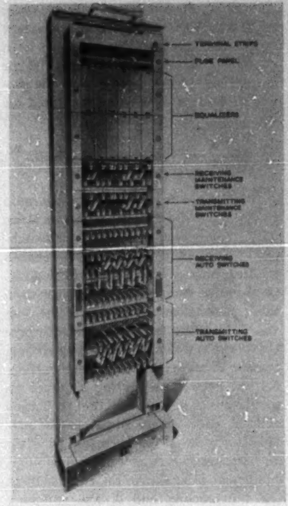

TERMINAL STRIPS
FUSE PANEL
EQUALIZERS
RECEIVING MAINTENANCE SWITCHES
TRANSMITTING MAINTENANCE SWITCHES
RECEIVING AUTO SWITCHES
TRANSMITTING AUTO SWITCHES

Fig. 7. I-f switching bay

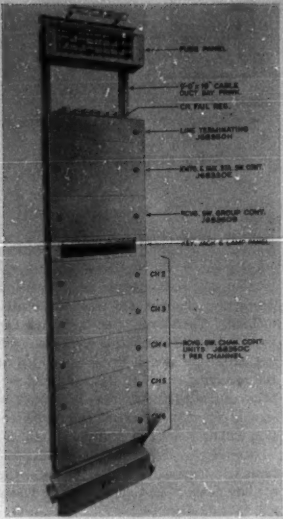

FUSE PANEL
9'-0" x 19" CABLE DUCT BAY FRAME
CR. FAIL REG.
LINE TERMINATING J68350H
XMTG. & MX. STA. SW. CONT. J68350E
RCVG. SW. GROUP CONT. J68350B
ALT. JACK & LAMP PANEL
CH 2
CH 3
CH 4
CH 5
CH 6
RCVG. SW. CHAN. CONT UNITS J68350C 1 PER CHANNEL

Fig. 8. Switching control bay

223-type coaxial switch, i-f bridging amplifier, regular and protection channel switch initiators, pilot supplies, and control circuits. Figs. 2 and 3 respectively show the schematic of the transmitting-end and receiving-end coaxial switches and bridging amplifiers which may be inserted in the transmitting and receiving i-f switch bays respectively of Fig. 1.

Coaxial Switch

The *223*-type coaxial switch shown in Fig. 4 performs the actual switching operations. The switch is designed for use on 75-ohm coaxial circuits carrying signals in the 70-mc range. The switch has four transfers which operate simultaneously when power is applied and crosstalk between circuits is 90 db or greater. The switching elements are glass-enclosed reeds which operate in 1.5 milliseconds. Positive contact is obtained by liquid mercury which wets the reeds by capillary action. The switch housing is a die-cast assembly and the power and coaxial connections are made by separable connectors.

I-F Bridging Amplifer

The i-f bridging amplifier is shown in Fig. 5. This amplifier has a single coaxial input and two coaxial outputs. The path from *A* to *B* is a low-pass filter with a cutoff above 400 mc, and transmission from 60 to 80 mc is flat. The path from either *A* or *B* to *AB* passes through a 1-tube amplifier which is usually adjusted for unity gain. Failure of the electron tube will not cause failure of the path *A* to *B*. Similar to the coaxial switch, the housing of the amplifier is die-cast and all connections are by means of separable connectors.

Regular and Protection Channel Initiators

The initiators contain the evaluating circuits. They accept a 70-mc signal from each radio channel and an 8.9-mc tone from a local signal generator. D-c control signals representing the condition of the received signal are passed to the control circuits. The initiator contains an f-m receiver in which the 8.9-mc tone amplitude-modulates the 70-mc carrier signal. The base-band output of the receiver contains the television or message information plus the 8.9-mc tone. If the carrier is lost, then the 8.9-mc tone is also lost to the output. The evaluating circuits analyze the 8.9-mc tone and the noise level in a narrow band around it. The protection channel initiator is similar to the regular channel initiator except that the evaluating circuits analyze the 8.5-mc tone and the narrow noise band about that frequency so as to determine the condition of the protection channel. The initiators are of conventional chassis construction for 19-inch-wide duct-type bays, and all of the initiators at the receiving end of a switching section are mounted in one bay, as shown in Fig. 6.

Pilot Supply

The 8.9-mc signal which is fed to the initiators is supplied by a pilot supply at the receiving end of each switching section. This unit contains two separate oscillators with a *223*-type switch and automatic transfer circuits to maintain output in the event that one oscillator fails. Construction is similar to the initiators.

Protection Pilot Resupply

This unit is used at the transmitting end of each switching section to supply the pilot information for the protection channel. Since the radio frequencies are not converted to base-band information at each repeater station and, hence, at the ends of some of the switching sections, this supply must provide the equivalent of an 8.5-mc tone frequency-modulated on a 70-mc carrier. Conventional mechanical construction is used similar to the other pilot supply equipment.

Filters

The automatic protection switching system depends upon analysis of the narrow band of information located about 8.5 and 8.9 mc in the base-band information. Although the input to the *TD-2* radio system is not normally expected to contain information at these frequencies, pilot-blocking filters are inserted in each radio channel at the input to the f-m terminal transmitters of the *TD-2* radio system to wipe this area of the band clean. These filters are of conventional filter design and are mounted with the terminal equipment.

Control Circuits

The control circuits are a combination of electromagnetic relay and electron tube circuits. The receiving-end control equipment generates the voice-frequency signals for use as guard tone, channel identification frequencies, and the 2,100-cps tone for the auxiliary station supplementary control signal. In addition, this receiving-end equipment times the switching operations, provides the recycling, and controls the receiving-end coaxial switches. The transmitting-end control equipment analyzes the voice-frequency signal received from the receiving-end equipment and controls the transmitting-end coaxial switches. This equipment also is of the conventional relay rack panel design.

Bay Arrangements

Although each direction of transmission of the *TD-2* radio system and the associated automatic protection switching is independent, 2-way service is provided over the *TD-2* radio system route and, hence, automatic protection switching is usually provided on a 2-way basis. Thus, at each end of the switching sections, transmitting-end equipment and receiving-end equipment generally will be provided. The initiators require an entire bay at each location, as shown in Fig. 6. However, the i-f switches can be mounted in a single bay, as shown in Fig. 7. Additional space in this bay may be used for other equipments which are not part of this protection system. The control circuits can be housed in one bay, as shown in Fig. 8.

Other Uses of Component Parts

The building blocks of this switching system have been adapted for other uses with the *TD-2* radio system. As indicated in Fig. 7, maintenance switching has been used to substitute other equipment on a maintenance basis. Program switching is another use.

Conclusion

Automatic protection switching for the *TD-2* radio system was put into service on the Denver and Sacramento route in June 1954. Since that time, other routes have been equipped with this protection. Tests and service observations have shown that the automatic protection switching system is providing marked improvement in the reliability of the *TD-2* radio system.

References

1. The TD-2 Microwave Radio Relay System, A. A. Roetken, K. D. Smith, R. W. Friis. *Bell System Technical Journal*, New York, N. Y., vol. 30, Oct. 1951, pp. 1041–77.

2. A Broad-Band Transcontinental Radio Relay System, T. J. Grieser, A. C. Peterson. *Electrical Engineering*, vol. 70, Sept. 1951, pp. 810–15.

3. An Unattended Broad-Band Microwave Repeater for the TD-2 Radio Relay System, R. W. Friis, K. D. Smith. *Ibid.*, Nov. 1951, pp. 976–81.

4. Protection of Service in the TD-2 Radio Relay System by Automatic Channel Switching, I. Welber, H. W. Evans, G. A. Pullis. *Ibid.*, vol. 34, May 1955, pp. 473–510.

An Instrument to Record Voltage Flicker

V. H. KRAYBILL
MEMBER AIEE

H. M. JENSEN
ASSOCIATE MEMBER AIEE

A VARIETY of electric devices have come into use in industry which impose rapid and frequent changes of load. This may cause sudden changes in voltage at frequent intervals, a condition which is known as voltage flicker. It is the concern of the electrical utility to detect such flicker to keep it below the level at which objectionable lamp flicker results.

It was found desirable to have an instrument which would give a continuous record of voltage changes. This could be connected to record the voltage at distribution substations or at any suitable location. If at any time the voltage flicker exceeded permissible values, an investigation could be made to determine the cause of the flicker and remedial action could be taken.

General Description

Certain difficulties are encountered in attempting to measure flicker voltage with conventional recording instruments. Since the voltage changes are very rapid and of short duration, most recording meters will not respond fast enough to follow these changes. Even though an instrument was available which would follow these variations, and which had a scale with sufficient expansion to show them, it was not practical to obtain a continuous record of all voltage changes. These changes are likely to occur several times a minute, and a chart with a speed suitable for continuous operation would be solidly inked. It would then be impossible to distinguish the individual voltage changes. For example, a 1-volt drop in voltage level followed by an additional 1-volt drop a few seconds later would appear as a 2-volt drop. However, the resulting lamp flicker would be that of two separate voltage drops of 1 volt each.

In reference 1 more details of the problems involved in measuring voltage flicker are given, two devices which were designed for this purpose are described and further references to other methods which have been used are mentioned. These other means of measuring voltage flicker have been designed primarily for staged tests, in which a complete record of voltage variations have been obtained for a short interval of time with the use of oscillographs with relatively high-speed charts or cameras. The device described in this paper records on a standard graphic voltmeter with a chart speed suitable for continuous operation. It does not give a record of voltage level or of all voltage variations but only records sudden voltage drops exceeding

---

Paper 56-797, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 12, 1956; made available for printing May 25, 1956.

V. H. Kraybill is with the Commonwealth Edison Company, Chicago, Ill., and H. M. Jensen, formerly with Commonwealth Edison Company, Chicago, is with Pennsylvania State University, University Park, Pa.

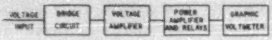


Fig. 1. Block diagram of the flicker recorder

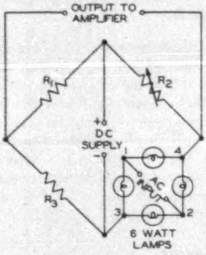


Fig. 2. Simplified diagram of the bridge circuit

certain predetermined values. Even though such voltage drops occurred rapidly enough to ink solidly the chart, the record still showed that the individual voltage drops had exceeded certain values. The exact number of such variations could not then be determined, but this is not generally required. A counter could be used in connection with the device to obtain that information.

The device was built to record three ranges of voltage flicker. For example, a drop exceeding 1.0 volt but less than 1.5 volts would move the pen to one-fourth of full scale on the chart. A voltage drop between 1.5 and 2.0 volts would give a deflection of one-half of full scale, and any drop exceeding 2.0 volts would record at three-fourths of full scale. By simple adjustments, these ranges can be set at any value from a fraction of 1 volt to several volts of flicker on a 120-volt base.

A time delay is provided which allows the use of the standard graphic voltmeter. Regardless of the duration of the voltage drop, the pen reaches the proper deflection and returns to zero before it responds to another voltage drop of the same or smaller magnitude. However, it is designed so that a voltage drop in a higher range will be recorded, even though it occurs while the pen is deflected for a lower value.

The response time of the recording meter limits the rate at which flicker voltages can be recorded. The limit of the device built is about one deflection per second. If the flicker occurs more often than this, the device continues to record at that rate. However, an output jack is provided on the device to which a higher speed oscillograph may be connected to record the exact rate of flicker. This record may be calibrated to obtain the actual value of the voltage drops rather than the voltage range normally recorded.

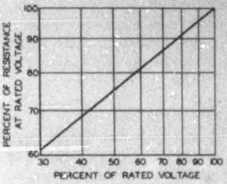


Fig. 3. Resistance characteristics of a 6-watt lamp with change in voltage

The measuring instrument itself does not have instantaneous response. There is a thermal lag introduced by the incandescent lamps used in the measuring circuit and an additional delay in the filtering circuit. In the present design, the over-all time of response is about 9 cycles (on a 60-cycle-per-second circuit). Voltage drops of shorter duration than this appear to be of less than the true value. If required, this response time could be reduced somewhat by a change in the circuit design.

An important consideration in designing the instrument was to use a measuring circuit which would respond to the drop in rms voltage rather than average or peak values. This is important when measuring flicker voltage produced by resistance welders or electric arc furnaces since the current is often nonsinusoidal and the voltage wave form may therefore be distorted. The rms voltage is desired because that is the value which determines the amount of lamp flicker.

Block Diagram

The components of the flicker recorder are shown by the block diagram of Fig. 1. The voltage being measured is connected to the bridge circuit through an isolating transformer. The nominal input voltage is 120 volts. The power requirement of this input is approximately 12 watts.

With a constant voltage input, the output of the bridge circuit is a direct voltage which is a function of the rms value of the input voltage. Changes in the input voltage will produce a varying direct voltage output. This voltage is applied to a 2-stage resistance-capacitance-coupled amplifier. This circuit is designed so that only the rapid variations in the voltage are amplified. If the input voltage is constant, the amplifier output is zero. The input voltage level affects only the grid bias of the first stage.

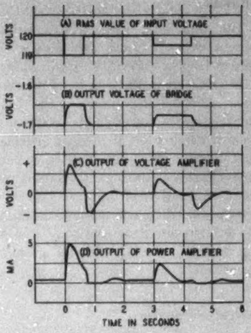


Fig. 4. Voltage and current changes at certain locations in the measuring and amplifying circuits when a change occurs in the voltage being measured

The output of the voltage amplifier is connected to a power amplifier which operates three sensitive relays. These relays may be adjusted to pick up at the desired values of voltage flicker. For example, one relay may operate with a 1.0-volt drop, with the other relays set for higher values. The relay contacts control auxiliary relays which operate the graphic voltmeter, the deflection depending on whether one, two, or three of the relays pick up. Therefore, a record is obtained which shows that the voltage flicker exceeded certain predetermined values. The auxiliary relays seal in, allowing time for the recording pen to reach the proper deflection. A time-delay relay then operates to allow the pen to return to zero.

Circuit Description

BRIDGE CIRCUIT

The detection of voltage variation is accomplished by the bridge circuit as shown in simplified form in Fig. 2. This bridge detects the change in resistance of lamp filaments resulting from a change in the applied voltage. This resistance varies with the temperature of the filament and is, therefore, a measure of the rms value of the voltage. The curve of Fig. 3 shows the relationship between the resistance and the voltage for the lamps used in the bridge. With a logarithmic voltage scale, this is a straight line over the range shown. Therefore, the percentage change in resist-

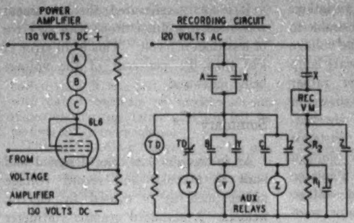
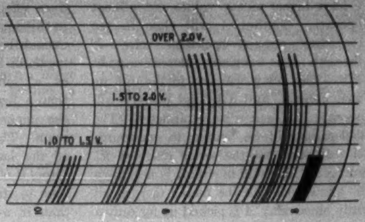


Fig. 5 (left). Simplified diagram of the power amplifier and recording circuit

Fig. 6 (right). Record obtained on graphic voltmeter



ance will be the same for a given percentage change in voltage at any point on the curve. The operating point varies with the voltage level but this does not affect the calibration since the instrument is used only to detect rapid voltage changes. Aging of the lamps may shift the position of the curve slightly but, if the slope remains the same, the calibration will not be affected. Since the lamps are operated at only half voltage, there is very little aging effect and the life of the lamp is practically indefinite.

As shown in Fig. 2, the circuit consists of two bridges with the output terminals of the first bridge connected as one arm of the second bridge. The voltage to be measured is applied to terminals 1 and 2 of the first bridge which consists of four 6-watt incandescent lamps. Since the four arms of this bridge are identical, it remains balanced at any voltage. The purpose of the bridge connection is to balance out the alternating voltage at terminals 3 and 4. At these terminals, the bridge appears as a resistance which varies with the alternating voltage applied to terminals 1 and 2.

The other three arms of the second bridge are wire-wound resistors. A direct voltage is applied to this bridge, which is operated slightly unbalanced. Resistor *R2* may be adjusted to obtain the desired voltage output for a particular value of alternating-voltage input to the first bridge. Changes in this input will produce a change in the output of the second bridge. The final result is therefore a direct-voltage output which varies with the rms value of the alternating-voltage input.

In the actual circuit, a filtering capacitor is connected across terminals 3 and 4. A small 60-cycle voltage may otherwise appear across these terminals because of slight differences in the resistance of the four lamps. A 120-cycle component is also produced as the resistance changes during each half-cycle of alternating voltage. Additional filtering is provided at the output of the second bridge and in the following amplifier.

There is a component of direct current in the lamp circuit but its heating effect is small compared with the a-c component. The d-c supply is regulated to eliminate voltage variations from this source which would appear in the output of the bridge.

VOLTAGE AMPLIFIER

The output of the bridge circuit is connected to a voltage amplifier. A 2-stage resistance-capacitance-coupled amplifier is used. The output of the amplifier is a voltage pulse which occurs with each rapid change in voltage input. With a constant voltage input, the output voltage is zero.

The curves of Fig. 4 show the voltages which appear at various points in the circuit when a voltage change occurs. Curve *A* shows the rms value of the voltage input to the bridge circuit. The voltage drops shown are typical of those produced by an electric welder. The output of the bridge circuit is shown by curve *B*. The voltage changes are not instantaneous because of the thermal delay of the lamps and a delay in the filtering circuit.

The circuit parameters of the voltage amplifier are chosen to obtain a satisfactory response to this signal. The output voltage of the amplifier is shown by curve *C*. A voltage drop produces a positive pulse and a voltage rise results in a negative pulse. This signal is applied to a power amplifier.

POWER AMPLIFIER AND RECORDING CIRCUIT

A simplified diagram of the power amplifier and recording circuit is shown in Fig. 5. The power amplifier operates three sensitive relays, shown as *A*, *B*, and *C*. The pickup current of each relay may be adjusted with shunting resistors not shown in the diagram. Relay *A* is set for the lowest pickup value.

The output current of the power amplifier is shown by curve *D* of Fig. 4. A voltage drop will give a positive current pulse as shown. Since this tube is biased nearly to cutoff, the output current is zero for even a small negative signal. Therefore, only voltage drops are used to operate the relays.

The contacts of relays *A*, *B*, and *C* are used to operate auxiliary relays *X*, *Y*, and *Z* in the recording circuit. These, in turn, control the deflection of the recording voltmeter. For example, assume that a voltage drop of 1 volt occurs, as shown by curve *A* of Fig. 4. If the pickup value of relay *A* is adjusted to 5 milliamperes, it will just operate with this voltage drop. Relays *B* and

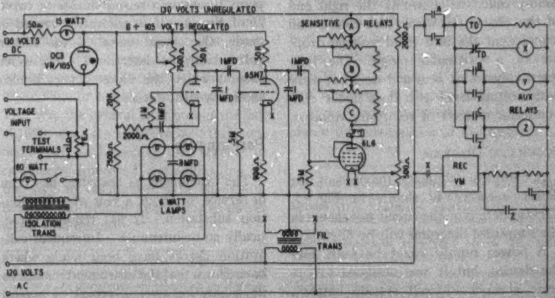


Fig. 7. Schematic wiring diagram of flicker recorder

C have higher pickup values and will not operate. Contact *A* will close in the recording circuit, operating relay *X* through the normally closed contact *TD*. This relay will seal in through its own contact, and a second contact will operate the recording voltmeter. The deflection will be limited to one-fourth of full scale by the series resistors *R1* and *R2*. The seal-in circuit is necessary to allow time for the meter to respond fully since relay *A* is picked up for only a few cycles. The coil of a thermal time-delay relay is also energized by the seal-in contact *X*. After a time delay, contacts *TD* open, dropping out relay *X* and de-energizing the recording meter. The circuit is then ready to record another voltage drop.

Assume that relay *B* is adjusted to pick up with a voltage drop of 1.5 volts and that relay *C* will pick up with a 2.0-volt drop. With any value between 1.5 and 2.0 volts, both relay *A* and relay *B* will pick up. The circuit will operate as described in the foregoing and, in addition, contact *B* will close to pick up relay *Y* which seals in. Resistor *R1* is by-passed by contact *Y*, allowing the voltmeter to reach one-half of full scale. If the voltage drop is over 2 volts, relay *C* will also pick up, operating relay *Z* which by-passes resistor *R2*. The meter deflection is then three-fourths of full scale.

A voltage drop may occur during the time the relays are sealed in and the meter is deflected. If this voltage drop is in a higher range than that being recorded, the meter will be deflected to the higher value. A voltage drop in the same or lower range will have no effect.

A reproduction of the record obtained on the recording meter is shown in Fig. 6. Three voltage ranges are shown for which the instrument might be calibrated. If the voltage flicker is occurring rapidly, the record will be solidly inked, as shown at the right end of the chart. However, the record still shows that the individual voltage drops are in a particular range. The base line is always at zero and does not vary with voltage level as it does in the conventional method of recording voltages.

SCHEMATIC DIAGRAM

A complete schematic diagram of the flicker recorder is shown in Fig. 7. Certain details of the circuit not shown in the simplified diagrams will be discussed.

A power supply could be built into the device, but it was designed to operate from a 130-volt station battery. A voltage-regulating tube is used to reduce effectively the voltage variations in this source. It was found desirable to extend the operating range of the voltage regulator by using a small 15-watt lamp as a voltage-dropping resistor rather than a fixed resistor. This is shown in the upper left-hand corner of the diagram. The decrease in the resistance of the lamp filament, as the voltage across it is reduced, permits the regulator tube to start at a lower supply voltage than it would if a fixed resistor were used of the resistance value required to limit the tube current to its rated maximum value when the voltage supply is at its maximum.

The voltage to be measured is connected to the terminals labeled "voltage input." A calibrating circuit is provided which may be used to produce the desired voltage drop. With a 60-watt lamp as shown, any drop between zero and 3 volts may be obtained by adjusting the rheostat to the proper value and operating the push button switch. If the source of voltage to be measured has insufficient capacity, any other 120-volt a-c source may be used for calibrating.

The supply voltage to the power amplifier is unregulated. However, the circuit design is such that variations in this voltage have a negligible effect on the output current. The effect of a change of plate voltage is compensated by the change of bias voltage which is obtained from the same source through a voltage divider.

The relays are calibrated by changing the bias of the power amplifier tube and by adjusting the shunting resistors across relays *B* and *C*. These resistors are arranged in a network designed to reduce the effect of an adjustment of one relay on the other relays.

An output jack is provided in the plate circuit of the power amplifier tube. An oscillograph may be connected at this point to obtain a record similar to curve *D* of Fig. 4. If a calibration record is made, the actual value of the flicker voltage may be determined. The exact rate at which the flicker occurs is also shown by such a record.

Performance

The instrument as described has been in service for about a year at a distribution substation. It has required practically no maintenance. Infrequent calibration checks have been made which have shown that the instrument maintains its calibration over a period of several months without adjustment.

If properly calibrated, the instrument will measure voltage drops with an error of less than 0.1 volt. This is an error of less than 0.1 per cent of the voltage being measured.

Summary

An instrument has been designed and built to measure and record voltage flicker such as is caused by electric welders. A standard graphic voltmeter is used to record voltage drops which exceed certain predetermined values. A time-delay circuit permits recording voltage flicker which occurs too rapidly for a graphic voltmeter to respond fully.

The instrument responds to the rms value of voltage. The graphic voltmeter records three ranges of voltage flicker. The actual value of the voltage flicker may be measured by connecting an oscillograph to the instrument. This will also give the exact rate at which the flicker is occurring.

References

1. ELECTRIC ARC AND RESISTANCE WELDING—III. *AIEE Special Publication S46*. "New Devices for Measurement of Welder Voltage Drops on a Power System," **Marvin Fisher, Jr., V. H. Kraybill, R. E. Young**. Oct. 1952, pp. 85–100.

◆

Discussion

O. Ramsaur (Pennsylvania Power and Light Company, Allentown, Pa.): The instrument described can be used as a simple means of continuously monitoring the voltage at a given location to detect the presence of sudden voltage drops and to record the approximate magnitudes and times of occurrence of noticeable drops. This should be quite adequate in many cases and will determine whether further information is needed in any case.

The popular curves of flicker tolerance are based on magnitude of voltage change and on the frequency of occurrence. It would be interesting to know how the several magnitudes of dips and their frequency are used in deciding whether voltage fluctuations are tolerable or intolerable. This question has bothered us for a long time.

Since the instrument records only sudden drops in voltage, it is probably assumed that for every sudden drop there will be a corresponding sudden rise. This would hold for fixed loads like electric welders but would not hold for spasmodic loads like arc furnaces. Our experience indicates that a sudden rise in voltage is just as annoying as a sudden drop of the same amount. It is noted that the term voltage "change" is used in several places in the paper where "drop" is meant.

It is noted that a 60-watt lamp is used as a calibrating load. It would seem to me that the inrush current of the lamp would lead to considerable errors because: 1. The recorder operates during the period before stable voltage change is reached and 2. for frequent operation, the inrush current will depend upon how long the lamp filament has cooled since it was last used. Experience with circuit breakers on model railroads indicates that the latter effect is felt for periods of 2 or more minutes.

The use of an a-c bridge as one leg of a d-c bridge is both clever and new to me. It avoids some of the difficulties in other methods of measuring small changes in voltage.

F. W. Warburton (New England Power Service Company, Worcester, Mass.): The authors have added a new and ingenious instrument to assist in a solution of the flicker-voltage problem. Within the range of this instrument, to those needing a record of the time that voltage dips occur, as well as their number and duration, the device should be of considerable assistance.

In my experience the major phase of the problem has been concerned with spot welders having time deviation from 1.5 to 4.5 cycles. Some multicylinder compressors have caused six to eight voltage dips per second, a region of maximum irritation. These dips were mostly 0.5 to 1.0 volts on a 120-volt basis. Unfortunately, the orders of brevity and frequency involved are out of range of the authors' instrument.

It would be interesting to know the error in voltage drops of shorter duration than nine cycles when the described device is operated on a 60-cycle system, and how the authors propose to shorten this 9-cycle limitation.

Southwest Oregon Radio System for the Bonneville Power Administration

D. J. MARIHART
ASSOCIATE MEMBER AIEE

D. L. WYLIE
MEMBER AIEE

THE southwest coast line of Oregon is rugged, primitive, and beautiful. From Gold Beach, near the Oregon-California border, the towns of Port Orford, Bandon, Coos Bay, Reedsport, and Mapleton lie on the sea at the foot of the coast range in a northward sweep of 144 miles.

A Bonneville Power Administration (BPA) 115-kv line serves this area. Starting at Eugene, the line penetrates the coastal range via the Siuslaw River valley, reaching the coast near Florence. From this point, it runs southward, paralleling U. S. Highway *101* and reaching a southern terminus at Gold Beach. Intermediate substations are served at McKinley and Norway in addition to stations located at the towns previously mentioned, see Fig. 1.

The coastal range, in the 300-mile section south of Eugene, is traversed by only three main roads. Rain forest covers peaks ranging in elevation from 1,500 to 3,600 feet from bases at sea level. Rainfall on the western slopes averages 75 inches per year. For several months in the winter, snow up to 18 inches is experienced at elevations above 2,000 feet.

Important rivers flowing seaward are the Siuslaw at Florence, the Umpqua at Reedsport, the Coquille at Bandon, the Sixes at Sixes, and the Rogue at Gold Beach. All flood in the spring. The commerce of the area depends mainly upon the growing logging and forest product industry and ocean fisheries.

The BPA line to this area is the only large-scale source of power and the region is dependent upon it. Each year, spring gales with gust velocities up to 120 miles per hour take their toll of the area and the power line. Concurrent with the storms, heavy rainfall brings rivers to flood stage. Washouts of Highway *101* and branches are numerous. It is not uncommon for the area between Coos Bay and Gold Beach to be isolated and without electric power for periods of 1 to 3 days. Extensive losses to property and losses of life are experienced.

The job of prompt restoration of power lines, in the past, was complicated by the lack of communications following storms. Telephone service to this area was by wire lines. These usually went down in the early phases of the storms. With lines down, several radio amateurs in Coos Bay and vicinity were able to make sporadic contact with Eugene and other points. BPA line trucks equipped with medium-frequency (mf) radio sets operating on 2,222 and 4,880 kc were able to make some contact with each other and with Eugene while en route to the coast. Often the first job was to open Highway *36* down the Siuslaw valley and to remove many large fir trees from across the road. This was the general situation in the fall of 1950, when it was determined that, as a matter of public safety and service to customers, improved communication to the area in connection with the maintenance and prompt restoration of power was a vital necessity.*

Communications Requirements

It was determined that a radio network was needed to meet operating conditions. Communications requirements of the network are as follows:

IMMEDIATE

1. A 2-way party-line voice-conversation circuit between J. P. Alvey load-dispatching headquarters, Eugene, Oreg., and each of the coastal substations.
2. Very-high-frequency (vhf) and mf radio systems to provide coverage pattern over all lines and substations in the area. The design is to permit radio-equipped maintenance vehicles in the area to maintain continuous 2-way conversation with the J. P. Alvey dispatcher.

FUTURE

In line with projected growth of the area and the resultant increase in load demand, it was also necessary to consider the following probable future requirements:

1. Need for a second voice channel between J. P. Alvey substation and the southwest Oregon maintenance office at Coos Bay.
2. One or more telemetering channels between the aforementioned points.

While the future requirements were not controlling in the initial phase, there appeared a high probability that one or both would become a necessity within 5 years following energization. Consequently the planning of the entire system was tempered by these considerations.

A team was immediately organized to check the possibilities of utilizing one of

---

* The West Coast Telephone Company in cooperation with the Pacific Telephone and Telegraph Company has since improved telephone service considerably by placing in operation in October 1954 a 900-megacycle (mc) ultrahigh-frequency (uhf) system between Eugene, Florence, and Coos Bay.[1]

Paper 56-700, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 28, 1956; made available for printing April 30, 1956.

D. J. MARIHART is with the Bonneville Power Administration, Portland, Oreg., and D. L. WYLIE, formerly with the Bonneville Power Administration, is now with the Chicago, Milwaukee, St. Paul and Pacific Rail Road Company, Chicago, Ill.

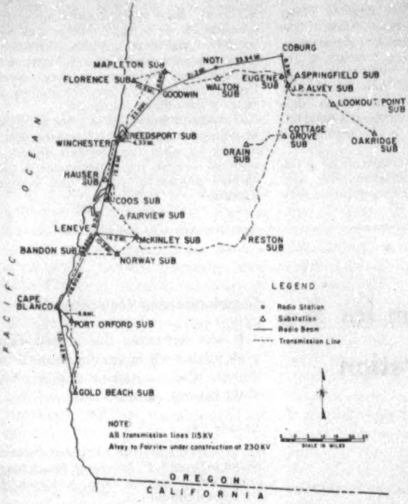


Fig. 1. Southwest Oregon radio system, location map

two available systems of communication. Attention was given first to possible expansion of the BPA mf radio network operating in the 2,222/4,880-kc band. Since the experience with this band had shown it to be generally unreliable during mid-day, attention was also directed to the possibility of a new vhf system operating in the 162- to 174-mc government band. This latter band has now become the standard for the BPA system replacing progressively the mf operation.

Three radio-equipped cars were used, one car was equipped with mf radio and two with 25-watt vhf sets. Directional corner-reflector type of antennas and portable masts supplemented the vhf equipment.

Office work prior to the survey consisted of checking U. S. Geodetic Survey quadrangle maps and BPA power facility maps to determine suitable routes and vhf repeater stations for the vhf system. It was known from the nature of the terrain that repeater stations would be necessary. Average airline distance between substations was 30 miles, and the substations were all at low elevation with surrounding land masks of varying magnitude. As has been previously noted, the problem of providing good vhf communications to the substations and along the lines was complicated by their separation from J. P. Alvey substation by the coast range.

Three separate routes were considered over the mountains. Of necessity, these were chosen to follow the highways over the mountains and are listed in the following in order of preference:

1. The Siuslaw Valley route via Coburg, Noti, and Mapleton; the route ultimately adopted is shown in Fig. 1.
2. The Umpqua Valley route via Drain, Elkton, and Reedsport.
3. The Bone Mountain route via Roseberg, Bone Mountain, and Norway.

Routes 2 and 3 were of interest as future planning called for new transmission lines in this area.

Profiles were drawn between prospective repeater stations on each of the three different routes. A typical profile is shown in Fig. 2. Stations were selected from the standpoint of available access roads, nearness to electric power (a difficult problem in this area), availability of land, minimum site elevation (the determining factor in winter being snow cover of site and access roads), and proximity to BPA substations and transmission lines which are necessary for good vhf coverage. The preliminary surveys were completed in 1 week.

Vhf signal strength readings were taken in terms of microvolts measured at the receiver input. Checks were made between each of the substations and each of the proposed repeater station sites. The preliminary data indicated that: 1. mf radio coverage could be obtained on a part-time basis between McKinley substation, Coos Bay substation, and J. P. Alvey; and 2. an economical vhf system could be installed with a maximum of five repeater stations. These stations would provide communications to all the BPA substations in question, and under all the 115-kv lines in the area for mobile-equipped line-maintenance trucks. Since the mf system could not provide coverage for all stations, it was eliminated from future consideration.

At this juncture it was realized that design factors predicted a dual system. A uhf or superhigh-frequency (shf) point-to-point backbone would be required, controlling vhf radio stations at the repeater and terminal points, for area coverage of the lines and substations. Available government band frequencies for the point-to-point system were: band *a*, 406 to 420 mc; band *b* 1,700 to 1,850 mc; and band *c*, 6,525 to 7,525 mc.

Single-channel equipment was known to be available operating in band *a*. Multichannel equipment was available from several manufacturers for operation in bands *b* and *c*.[2,3,4] A report was compiled, complete with cost estimates recommending a combination microwave and vhf radio system. Fig. 3 shows a block diagram of the system as constructed.

Point-to-Point Microwave Links

The southwest area dispatching headquarters is located at J. P. Alvey substation and is connected with J. D. Ross substation near Portland by a Federal Telephone and Radio-Corporation type-*10C* microwave system. It was decided to start the radio system to the coast from the Coburg microwave site, with the use of one of the 23 voice channels between Alvey and Coburg as the first link in the system. From Coburg, additional shf links were established as follows:

	Path Length, Miles
Coburg-Noti lookout	23.4
Noti lookout to Goodwin lookout	21.2
Goodwin lookout to Winchester	23.9
Winchester to Leneve	31.2
Leneve to Cape Blauco	28.9

Vhf links were used to extend service from the repeater aforementioned points to the various substations. As the microwave system ends at Cape Blanco, the final extension to Gold Beach substation, the southern terminus of the system, is by vhf. Since the communications conditions differ between individual re-

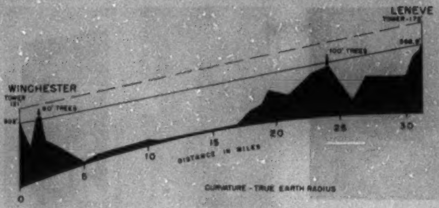


Fig. 2 (left). Profile of Leneve–Winchester beam path

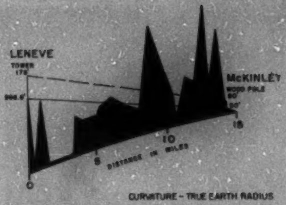


Fig. 4 (right). Profile of Leneve–McKinley beam path

peaters and substations, a brief résumé will be given for each point.

1. Noti: A vhf land station is arranged for continuous omnidirectional transmission to mobiles operating between Eugene and Greenleaf on Highway *36*.

2. Goodwin: A vhf land station is arranged for continuous omnidirectional transmission to substations and mobile units. It provides communications to Mapleton and Florence substations as well as highway coverage from Greenleaf to Winchester Bay. Vhf radio equipment at Mapleton and Florence substations is arranged for push-to-talk half-duplex directional transmission to the shf station.

3. Winchester: A uhf land station coverage of Highway *101* from Florence to Coos Bay. The service to Reedsport and Coos substations is on the same basis as in point 2.

4. Leneve: A vhf land station coverage of Highway *101* from Winchester Bay to Bandon. It serves the McKinley and Norway substations on a nonline-of-sight path on the same basis as in point 2. The service to Bandon is the same as for Norway and McKinley, except that the beam path is line of sight; see Fig. 4.

5. Cape Blanco: A vhf coverage from Bandon to Gold Beach. The service to Port Orford and Gold Beach substations is on the same basis as in point 2.

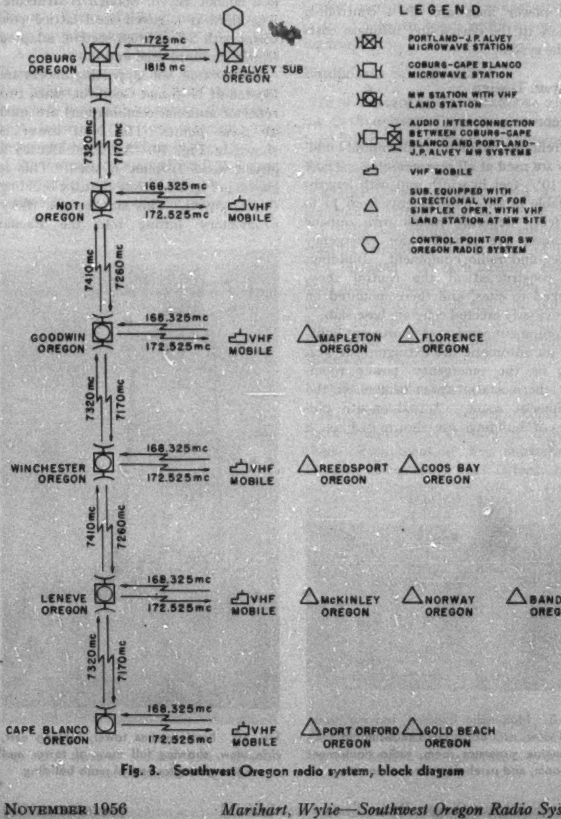

LEGEND
PORTLAND–J.P. ALVEY MICROWAVE STATION
COBURG–CAPE BLANCO MICROWAVE STATION
MW STATION WITH VHF LAND STATION
AUDIO INTERCONNECTION BETWEEN COBURG–CAPE BLANCO AND PORTLAND–J.P. ALVEY MW SYSTEMS
VHF MOBILE
SUB. EQUIPPED WITH DIRECTIONAL VHF FOR SIMPLEX OPER. WITH VHF LAND STATION AT MW SITE
CONTROL POINT FOR SW OREGON RADIO SYSTEM
COBURG OREGON
1725mc
1815mc
J.P. ALVEY SUB OREGON
7320mc
7170mc
NOTI OREGON
168.325mc
172.525mc
VHF MOBILE
7410mc
7260mc
GOODWIN OREGON
168.325mc
172.525mc
VHF MOBILE
MAPLETON OREGON
FLORENCE OREGON
7320mc
7170mc
WINCHESTER OREGON
168.325mc
172.525mc
VHF MOBILE
REEDSPORT OREGON
COOS BAY OREGON
7410mc
7260mc
LENEVE OREGON
168.325mc
172.525mc
VHF MOBILE
McKINLEY OREGON
NORWAY OREGON
BANDON OREGON
7320mc
7170mc
CAPE BLANCO OREGON
168.325mc
172.525mc
VHF MOBILE
PORT ORFORD OREGON
GOLD BEACH OREGON

Fig. 3. Southwest Oregon radio system, block diagram

A temporary mf station was installed at McKinley Substation, an interchange point with the California-Oregon Power Company, to provide interim communication until the vhf system could be installed. The project in entirety was approved in May 1952, with system energization scheduled for January 1, 1954. The work involved in providing this communication system can be divided into four main categories:

1. Field surveys.
2. System design.
3. Equipment specifications.
4. System installation and operation.

Field Surveys

Surveys were made to determine the following:

1. Microwave point-to-point path clearances.

2. Vhf field strength along roads, lines, and on nonline-of-sight paths.

3. Station elevations, and elevations of critical points.

4. Location of station sites, access roads and power lines.

5. Co-ordination data for the joint occupancy of U. S. Forest Service lookouts at Goodwin and Noti.

Microwave paths were checked with the use of heliographs in daylight and high-power lamp techniques at night. Terminal and intercept points were triangulated in the field and cross-checked in the office with data from a Kelsh plotter as derived from aerial stereo pairs.

Vhf field strengths on nonline-of-sight paths were established with the use of a 25-watt base station with a 36-foot portable mast and a quarter-wave whip antenna;

Fig. 5. Portable vhf test station

see portable vhf test equipment, Figs. 5 and 6. At the receiving end, a corner reflector and calibrated receiver were employed. On the intercept paths to Mapleton, McKinley, Norway, and Port Orford, reception was from waves diffracted over intervening ridges.[5] Careful check was made to determine the values of nulls and maximums in the standing wave pattern.

Access roads to microwave sites varied in length and condition. At Winchester, 1,200 feet of new road were needed. Twelve miles of existing U. S. Forest Service road were used at Goodwin with some crushed rock improvement of soft spots. An existing 3-mile access road was used at Cape Blanco, the site of a U. S. Coast Guard Loran station. The balance of the sites were served by existing county roads or highway *101*.

Power lines were a more difficult problem. It was necessary to contract with local utilities for power service extensions to repeater sites as follows:

Noti lookout: Construct 2 miles of new line.

Goodwin Peak: Construct 7 miles of new line.

Winchester Bay: Construct 600 feet of new line.

Leneve: Construct 2 miles of new line.

Fig. 6. Close-up of test equipment in Mobile

Fig. 7. Power line to Goodwin Peak

The line construction to Goodwin Peak was especially difficult, as can be seen from Fig. 7 which shows one span of the line.

At all other sites, power at distribution voltages was available. Costs for initial construction of the lines listed in the foregoing were substantial and proved that in virgin terrain the costs of roads and power lines are often controlling factors in the design and ultimate costs of the system.

System Design

BUILDINGS

Prefabricated modular aluminum buildings are used at all repeater sites. These are $10^{1}/_{2}$ feet in width and with lengths variable in 2-foot increments from 14 to 20 feet. Separate rooms with outside entrances are provided for emergency power and radio equipment. Buildings are constructed at the central shop, trucked to sites, and there mounted on a previously erected concrete base slab.

Equipment consists of fluorescent lighting, an automatic fire extinguishing system in the emergency power room, and thermostatic space heaters in the equipment room. Actual on-site pictures of buildings are shown in Figs. 8 and 9.

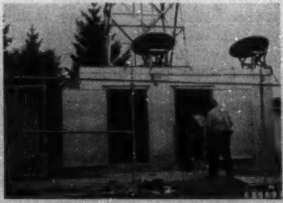


Fig. 8. Noti radio building, looking north, front view, with doors open showing close-up of engine generator room, radio equipment room, and parabolic antenna mounting

Fig. 9. Noti radio building, inside view showing close-up of engine generator equipment on left and radio equipment on right

TOWERS

Three general types are employed. The first is a 50-foot self-supported lightweight-steel transmission type with top-mounted platform. The second is a 50-foot guyed, wooden H-structure. The third is a guyed steel lattice pole tower with 2-foot cross section, adapted from a standard dead-end tower.

The 50-foot self-supporting towers are located at Noti and Goodwin, since two reflector antenna combinations are used at these points. The Noti tower is shown in Fig. 10. At Cape Blanco a guyed wood H-frame is used. This is shown in Fig. 11. Note that the building is also guyed. This site is on a rocky promontory jutting into the Pacific

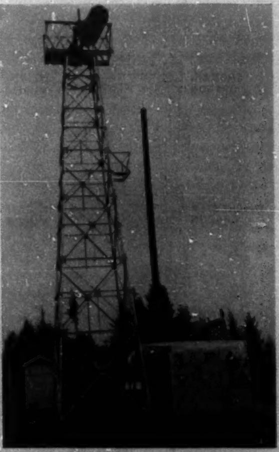


Fig. 10. Noti antenna tower, looking east, side view, showing full view of tower and engine generator and of radio building

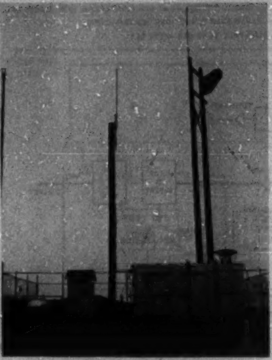
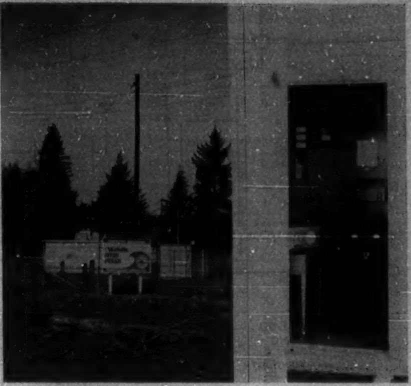


Fig. 11. Cape Blanco, looking west. Front view with door closed showing guys

Fig. 13. Reedsport substation, looking north. Front view of control house with doors open. Control console is shown on table to left of vhf radio panel



Ocean. Winds of 120 miles per hour in gusts have been recorded.

A 120-foot guyed lattice pole is used at Winchester and a similar tower of 175-foot height at Leneve. Use of these towers made it possible to span a close intercept between the two stations. The tower at Leneve is shown in Fig. 12.

All towers are designed for 0.25-degree torsional deflection and 0.5-degree vertical deflection under maximum wind and ice load conditions. The 4- by 6-foot reflectors used on all towers have an approximate 3.5-degree beam width at operating frequency.

Equipment Specifications

In accordance with specifications covering the purchase of system components, the following basic system was obtained:

EMERGENCY POWER

Kohler 5-kva propane-driven engine generators, single phase, 60 cycles, 117 volts a-c, equipped with automatic starting, stopping, and transfer of load were installed. At each location, except Noti and Winchester, 200-gallon tanks are used which provide sufficient reserve for a 10-day operation. At Noti and Winchester 500-gallon tanks provide a 30-day reserve.

VHF RADIO INSTALLATION

Motorola 25-watt vehicular-type vhf radio equipment is installed at Substations. Equipment at five substations, Florence, Coos Bay, Port Orford, Norway, and Bandon, provides a 6-volt operation from a *200 AH* battery supply and charger. Balance of stations have 117-volt a-c equipment, with emergency supply from rotary converters operating on 24-volt batteries.

The 6-volt stations are unattended, the radio equipment being turned on only when the operator is at the station. Stations at Mapleton, McKinley, Reedsport, and Gold Beach have operators and use 117-volt equipment, which is equipped with annunciators so that the dispatcher may call the operator. The layout of Reedsport substation is shown in Fig. 13.

A 117-volt a-c 50-watt vhf base-station equipment is used at all microwave repeater points. Twelve BPA maintenance trucks were initially equipped with vhf mobile sets. A close-up of the control head and installation are shown in Fig. 14.

SPECIAL FEATURES

The vhf system incorporates the following unique features:

1. Base stations operate at full duplex at all times, using transmit frequency of 172.525 mc and receive frequency of

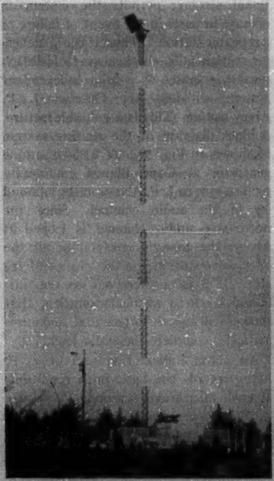


Fig. 12. Leneve, looking east. Over-all station picture

Fig. 14. Mobile installation showing control head

Fig. 15. Western Electric type-508A switchboard, J. P. Alvey

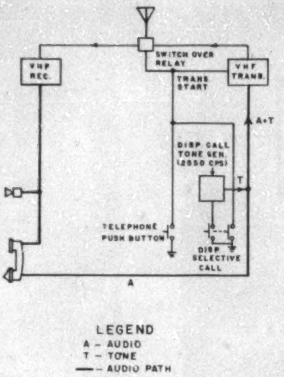

SWITCH OVER RELAY
VHF REC.
VHF TRANS.
TRANS. START
A+T
DISP. CALL TONE GEN. (2550 CPS)
T
TELEPHONE PUSH BUTTON
DISP. SELECTIVE CALL
A
LEGEND
A – AUDIO
T – TONE
— – AUDIO PATH

Fig. 16. Block diagram, mobile selective call

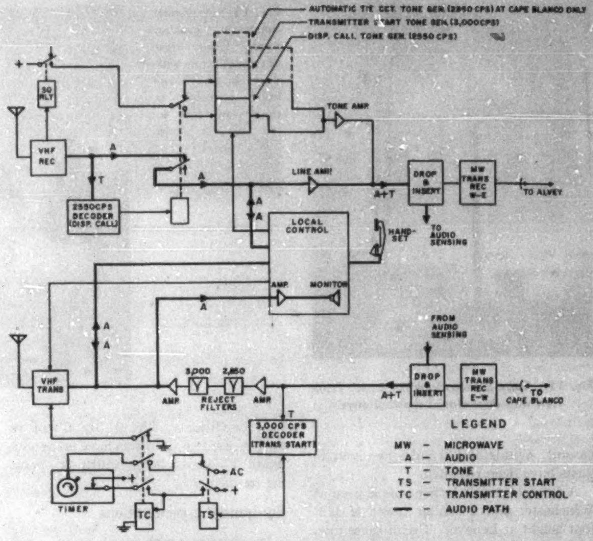

AUTOMATIC TIE CCT. TONE GEN. (2850 CPS) AT CAPE BLANCO ONLY
TRANSMITTER START TONE GEN. (3,000 CPS)
DISP. CALL TONE GEN. (2550 CPS)
+
SO RLY
VHF REC
A
T
2550CPS DECODER (DISP. CALL)
TONE AMP.
LINE AMP.
A+T
DROP & INSERT
MW TRANS REC W-E
TO ALVEY
TO AUDIO SENSING
LOCAL CONTROL
HAND-SET
AMP.
MONITOR
FROM AUDIO SENSING
VHF TRANS
AMP
3,000
2,850
REJECT FILTERS
AMP
A+T
DROP & INSERT
MW TRANS REC E-W
TO CAPE BLANCO
3,000 CPS DECODER (TRANS START)
AC
+
TIMER
TC
TS
LEGEND
MW – MICROWAVE
A – AUDIO
T – TONE
TS – TRANSMITTER START
TC – TRANSMITTER CONTROL
— – AUDIO PATH

Fig. 17. Block diagram, base station selective call

168.325 mc. Transmitting and receiving antennas are physically spaced both vertically and horizontally. With this spacing and 4-mc frequency separation, cavities are not required to prevent receiver desensitization.

2. Any substation or mobile unit may selectively call the dispatcher or the attendant switchboard operator at J. P. Alvey substation. The system terminates on a multiple basis on the dispatchers Western Electric *508A* switchboard and on the Western Electric type-*551* attendant private branch exchange. Calling lamp and buzzer are provided together with interchange of "busy" indication. A picture of the type-*508A* switchboard is shown in Fig. 15. Outbound calling on the system is voice-call, with all land stations energized, a requirement of the BPA branch of operation.

3. Mobiles by means of hand-set push-to-talk buttons may select the closest of several vhf base stations within range and cause it to connect automatically to the microwave link for in bound transmission to Alvey. At Alvey, the inbound channel is normally connected to the outbound channel so that any mobile using the link automatically rebroadcasts over the entire area. Under these conditions, the dispatcher can override to take control when desired. He is therefore able to take charge of the system and talk or monitor at all times. Block diagrams of the controls on mobiles and at base stations are shown in Figs. 16 and 17. When the mobile operator pushes the "dispatcher" selective call button, he starts the vhf transmitter and simultaneously modulates it with a 2,550-cycle-per-second (cps) audio tone. This audio tone is received at the vhf base station, decoded, and used to start another 2,550-cps tone generator which sends the tone to the dispatcher headquarters at J. P. Alvey. At J. P. Alvey it is decoded and used to operate the calling lamp and buzzer on the telephone switchboard. The dispatcher operates his answer key, thereby energizing a 3,000-cps tone generator which transmits its tone on the outbound leg of the microwave audio channel. The 3,000-cps tone is received at each vhf land station, decoded, and used to turn on all the vhf transmitters.

4. Mobiles are equipped with two transmitters and one receiver, normally transmitting on f_2 and receiving on f_1. They work with the base station which transmits on f_1 and receives on f_2. When direct mobile-to-mobile communications are desired, the operator may switch the car transmitter to frequency f_1. These transmissions do not interfere with transmissions between base stations and mobiles in other areas. Since receivers are always tuned to f_2, the mobile may be contacted by the dispatcher at any time by voice call.

MICROWAVE POINT-TO-POINT SYSTEM

The specifications permitted radio equipment suppliers to bid on either uhf or shf equipment. Philco *CLR-6* microwave radio equipment was low bid. Operation is in the 7,125- to 7,425-mc shf band, and provides the backbone system.

The audio base band is used. There being no special multiplexing equipment, the system operates with amplitude and frequency modulation. System fade margins are set at 20 decibels per hop (about 3/4 decibel per mile). *F1A* weighted signal-to-noise ratio for a single hop is about 50 decibels. Under idle channel conditions the outbound and inbound voice channels are connected at Alvey so that a true party-line system is provided.

An automatic audio tie, with directional sensing, permits the system to operate in parts in the event of failure of a repeater station. Thus, if the Winchester station fails, the Leneve–Gold Beach portion operates as a fully independent system, as does the Goodwin–J. P. Alvey section. This is a valuable feature. A block diagram of the sensing system is shown in Fig. 18. A 2,850-cps tone generator at Cape Blanco continually sends a tone to J. P. Alvey, on the inbound leg of the audio channel. Since the microwave audio channel is looped at Alvey, the tone is received at all the microwave stations on the outbound leg also. If it is not received on the outbound leg at a particular station, that station will switch to terminal, and automatically complete the audio loop.

An alarm indicator system at J. P. Alvey records the basic fault conditions at each microwave station. These are: 1. east-to-west radio-frequency failure; 2. west-to-east radio-frequency failure; and 3. emergency power plant operating. Inbound alarms are by sequential time-coded tones.

The alarm equipment is important in pin-pointing troubles since economic

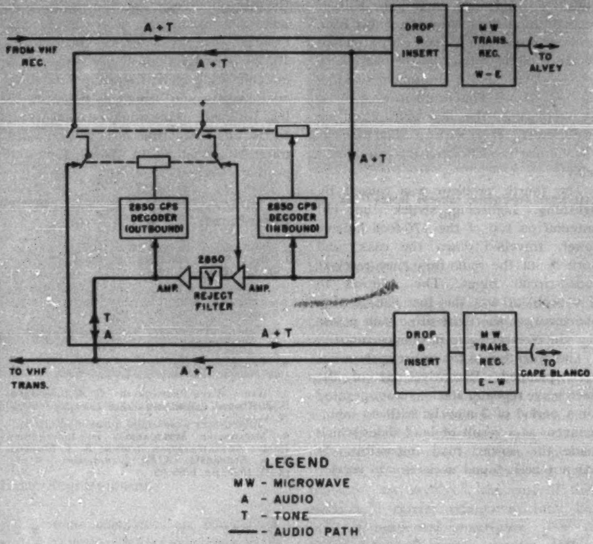


Fig. 18. Block diagram, audio sensing

Table I. Breakdown of Costs

Item	Per Cent of Total
Field surveys	7
Engineering	4
Land	1
Microwave equipment (including installation labor cost)	27
Vhf radio equipment (including installation labor cost)	
Base stations	11
Substations	8
Mobiles	7
Housing	5
Site development	13
Towers	2
Emergency power	3
Commercial power	8
Roads	1
Test and energization	3
Total	100

conditions did not permit radio-frequency stand-by equipment at the stations, although the system is designed to permit addition of stand-by equipment at a future date.

Components of System Cost

Table I represents the breakdown of costs, in percentages, on this particular job, and gives an indication of what may be expected in locations of this nature for systems of the type outlined. Of the items listed in Table I, engineering, radio equipment, housing, towers, emergency power, and test and energization are quantities that may be accurately estimated. Factors in which large cost uncertainties were found to exist in this case were the site development, commercial power, and roads. Tower costs were substantial but often cheaper than additional repeater stations. In this case, the high towers at Winchester and Leneve obviated the need for an intermediate station originally planned.

System Installation and Operation

The backbone system consisting of all the microwave links and two vhf links was installed in a record-breaking 6 months after it was placed on order. The remainder of the system consisting of seven vhf links was installed 18 months after the backbone system.

Several interesting problems were encountered during the installation period. The first involved a low fade margin on the Leneve-to-Cape Blanco microwave link. Approximately one-third of this link is over water. Since it was imperative that cost be held to a minimum on this particular system, no precautions were taken against unusual fades such as were known to exist on other over-water paths. During 2 months of operation, the microwave radio-frequency signal strength charts recorded fades greater than 25 decibels. A typical chart recording is shown in Fig. 19. This was serious and resulted in momentary system outages, since the fade margin for this link was set at 25 decibels. Steps were immediately taken to replace the 4-foot parabolic antenna and the 4-by-6-foot reflector at Leneve with a 6-foot parabolic antenna and a 6-by-8-foot reflector respectively. The larger

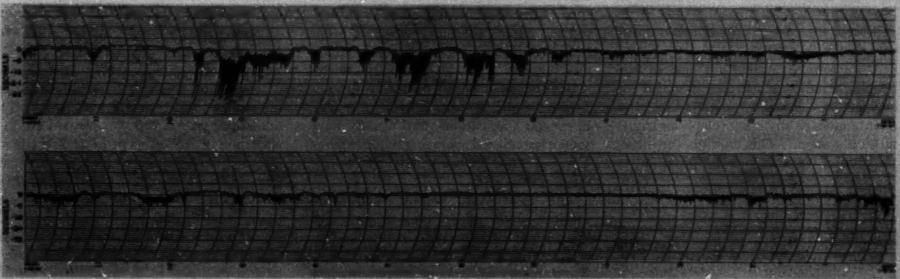


Fig. 19. Cape Blanco–Leneve microwave received signal chart recording

antenna and reflector provided an additional 8-decibel fade margin for a total of 33 decibels on this path. No momentary system outages caused by fading have been recorded or noticed since this change was made.

The second problem concerned the difference in phase at J. P. Alvey between dispatcher audio signal tones picked up at two different microwave stations. This problem was solved by adjusting the vhf receiver noise squelch sensitivity at the base stations to cut out weak signals and by reducing the power output of the Bandon vhf station so that the signal received at Cape Blanco was less than the vhf station squelch setting.

The third problem was discovered in November 1955 during a major power system outage in southwest Oregon. The power system frequency dropped to 57 cps at the microwave station but the voltage remained at 117 volts a-c. Since no frequency-sensitive relays were initially provided in the emergency power switch-over equipment, the engine generator set did not start. The voltage regulators on the microwave power feed, being of the resonant circuit type, caused the alternating voltage to the microwave to be too low for the equipment to function properly. This problem was solved by replacing the low-voltage-sensing switch-over relay with a dual-purpose low-voltage frequency-sensing switch-over relay.

The fourth problem was caused by lightning. Lightning struck the vhf antenna on top of the 175-foot Leneve tower, travelled down the coax, and burned out the radio-frequency receiver tuned-circuit input. The solution to this problem was the installation of a quarter-wave short-circuiting stub across the vhf receiver radio-frequency input.

The operation of the entire system has been planned. Recently, one of the microwave repeater stations was operated for a period of 3 months with no maintenance, as a result of land slides which made the service road impassible. It has not been found necessary to service the vhf equipment at the substations more often than once every 2 months. Over-all system in service time during 1955 was better than 97 per cent. It is possible that radio-frequency stand-by equipment would bring this figure up to 99.7 per cent. This would be comparable to the in service time achieved on BPA microwave systems in Washington and northern Oregon.

References

1. Microwave in the West Coast System L. Gray Beck. *Fortnightly Telephone Engineer*, Chicago, Ill., Dec. 1, 1954.

2. Microwave Channels for Power System Applications, AIEE Committee Report. *AIEE Transactions*, vol. 68, pt. I, 1949, pp. 40–43.

3. Radio at Ultra-High Frequencies. *RCA Review*, Princeton, N. J., vol. II, July 1948.

4. Frequency Modulation, Murray G. Crosby. *Ibid.*, vol. I, Jan. 1948, pp. 57–64.

5. Radio Wave Propagation, C. R. Burroughs, S. S. Atwood. Academic Press, Inc., New York, N. Y., 1939.

6. Microwave Applications to Bonneville Power Administration System, R. F. Stevens, T. W. Stringfield. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 1088–99.

Microwave for Telephone Companies

W. C. FISHER
MEMBER AIEE

MICROWAVE is considered very carefully before it finds itself in the toll plant of a telephone company. To an even greater extent than other communication services, telephone companies require microwave to be capable of the utmost in performance and reliability, and yet be economical. In fulfilling these obviously contradictory requirements, microwave must meet the unusual and exacting specifications of the telephone plant and, often at the same time, compete with the cost of time-honored open-wire and cable facilities.

A microwave system for use in the telephone industry must be designed for engineering, ordering, installation, and maintenance as any other telephone transmission system. This requires the incorporation of features and the provision of services which are often unfamiliar to the radio industry, but which are based on long-established telephone engineering practices.

Microwave systems range in size from small, packaged terminals that provide two or three channels to elaborate equipment capable of simultaneously carrying several color television channels and hundreds of telephone channels across the continent. They can be divided into three categories: heavy-route, medium-route, and thin-route systems. The heavy-route, or large-capacity systems, use Western Electric type-*TD-2* microwave, and are generally well known. The 1- to 3-channel systems are generally used in special applications, and usually have somewhat relaxed operating requirements. These systems are not widely used by the telephone industry.

The in-between system, known as medium-route microwave, is not well known, yet it is the type of facility that is most applicable to the average telephone plant. A typical telephone office installation of medium-route microwave equipment is shown in Fig. 1. There is no distinct division between medium- and heavy-route, or medium- and thin-route microwave; however, it is expected that the capacity of medium-route systems will eventually reach several hundred telephone channels. At least one system now in use can be expanded, in groups of 12 channels, to an ultimate capacity of more than 350 channels. At present, such medium-route microwave systems are being used by many telephone companies ranging from small Independents to Bell System Telephone operating companies.

Two frequency bands are most commonly used for common-carrier service. These are from 890 to 940 mc (megacycles), and from 5,925 to 6,245 mc. A distinct advantage of the former is that it allows the use of relatively inexpensive, flexible r-f (radio-frequency) transmission lines. This permits physical separation of the antenna and the microwave equipment and, thus, a versatility in regard to installation of the transmitter and receiver. The 900-mc equipment can always be installed indoors and, consequently, gets the advantage of routine maintenance at the system terminals. An advantage of the 6,000-mc band is the wider available bandwidth which permits larger numbers of channels to be transmitted without interference.

Paper 56-699, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 19, 1956; made available for printing May 15, 1956.

W. C. Fisher is with the Lenkurt Electric Company, Inc., San Carlos, Calif.

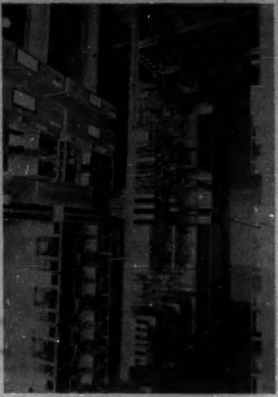


Fig. 1. Typical telephone office installation of medium-route microwave equipment

Basic Requirements

Telephone companies are usually operated by men with long experience in the techniques of communication by open-wire and cable. Consequently, to prove-in and be accepted, medium-route microwave must meet the majority of the following requirements:

1. Installation cost should be below that of an open-wire or cable system capable of providing the same number of channels. Installation costs include all costs incurred up to the line-up of the microwave system.
2. Quality of transmission should be equal to or exceed that of an equivalent open-wire or cable system.
3. Reliability and life expectancy should be comparable to that of other electronic equipment in the telephone toll plant, such as carrier telephone and telegraph equipment.
4. Microwave equipment must be mechanically and electrically compatible with other telephone equipment.
5. Maintenance cost should be equal to or less than that of an equivalent open-wire or cable system. Maintenance procedures should be simple enough to require little extra training of maintenance personnel.

A microwave manufacturer, in attempting to meet these five requirements, must realize that the provision of microwave equipment is only the beginning in fulfilling a telephone company's needs. Many other factors concerned with the installation, operation, and maintenance of the system take on as much significance as the characteristics of the basic radio equipment itself.

Fig. 2. Typical remote microwave repeater site

Installation Costs

Installation cost is one of the most important of the five basic requirements of microwave for the telephone industry. An important part of the total installation cost may be the cost of the microwave equipment. This is especially true for relatively simple installations. In many cases, however, particularly where large antenna structures, emergency power supplies, power lines, access roads, and buildings are involved, the cost of the radio and carrier equipment may be only of incidental importance. Fig. 2 shows a typical remote-repeater station in which the cost of the transmitting and receiving equipment was only a small part of the total installation cost. Fig. 3 shows a simple pipe-mounted antenna for a relatively inexpensive microwave installation.

Of the many items contributing to the initial cost of a microwave installation, several factors, which are under the control of the manufacturer, are important in keeping the first cost as low as possible. These factors include the design of equipment especially for telephone use, the construction of equipment that promotes easy installation, and the provision of information and instruction material for telephone personnel. Another factor reducing initial cost is the availability of a completely engineered microwave system instead of component parts which then require expensive engineering coordination. Also, accurate and up-to-date engineering and ordering information supplied by the manufacturer saves time in writing specifications and requisitions and assures complete delivery of the equipment.

An important purpose in keeping the first cost of a microwave system to a minimum is to improve its competitive position with respect to open-wire and cable facilities. At present, the prove-in distance of a single-section microwave system may be as short as 15 miles. For distances less than this, cable and cable carrier still appear to be the most appropriate facilities. In the future, however, improved and more economical microwave techniques may be expected to reduce the prove-in distance to less than 15 miles.

Fig. 3. A simple pipe-mounted antenna installation

An instance is known where a single-section 48-channel system was installed over a 17.5-mile path between two cities. At each end of the system, antennas were mounted on pipe masts on top of the telephone company office buildings. In engineering the system, a comparison was made between the cost of the 48-channel microwave and an equivalent cable installation. It was determined that cable would have cost about $4,500 more than the radio and multiplexing equipment installed and ready for service. In this particular case, the difference in cost was important; but the deciding factors were the expandability, reliability, and the ease and simplicity of installation of the microwave. Table I gives a comparison between the installation costs of an expensive and of a simple installation.

Transmission Quality and Performance

Medium-route microwave usually is expected to provide transmission facilities for part of, or all of, one link in the nation-wide toll network. It is therefore necessary for the microwave system to have characteristics which maintain satisfactory toll-quality performance.

The four most important transmission characteristics of a microwave system are flat frequency response, level stability, frequency stability, and adequate signal-to-noise ratio. The first three usually are the least troublesome to the microwave manufacturer because they represent areas in which conventional radio engineering techniques are applicable.

Table I. Comparison of Costs Between an Expensive and a Simple Installation

Expensive Repeater Installation, Dollars		Simple Terminal Installation, Dollars	
Land and building	25,000	Antenna support	500
Access road	18,000	Installation and accessory equipment	1,000
Antenna supports	6,000	Radio equipment†	8,000
Installation and accessory equipment	80,000	Total	9,500
Radio equipment*	20,000		
Total	149,000		

In an average installation, the microwave equipment cost amounts to less than 50% of the total installation first cost.

* Radio equipment cost is 13.5% of total cost.

† Radio equipment cost is 84% of total cost.

The operational characteristic which gets the closest scrutiny by telephone engineers is the ability of the microwave to provide a transmission medium in which all carrier channels are within the noise limits prescribed by accepted toll standards. The noise contribution of the microwave is the factor which limits, more than any other single characteristic, the system's length and channel capacity.

In the engineering of links (toll trunks) for the nation-wide toll dialing system, a frequently quoted objective is the maintenance of total long-term (more than 50% (per cent) of the time) average noise below +31 dba (decibels adjusted). The noise measurement quoted is at a zero-level transmission point with *F1A* weighting. The noise level per link may rise above this for small percentages of the time.

A link may be made up of combinations of cable, wire line, and microwave in tandem. Therefore, the microwave system's noise contribution will be limited to a value which is the difference between the accepted standard and the added noise power contribution by the cable, wire line, and channelizing equipment. Since each application of microwave involves different factors, it is necessary that the manufacturer provide accurate noise performance data to the telephone transmission engineer.

The noise contributed by the microwave system comes from three main sources:

1. The modulator of the transmitter (modulator noise).
2. Intermodulation distortion caused by nonlinearities in the microwave equipment.
3. The head end of the receiver (receiver thermal noise which is of principal interest during fading).

The importance of the noise contributions from these sources depends upon the number of repeater sections in a system, the section lengths, and the propagation path clearances. The noise produced by each source must be kept as small as design and manufacturing techniques will allow because individual noise contributions are additive on a power basis.

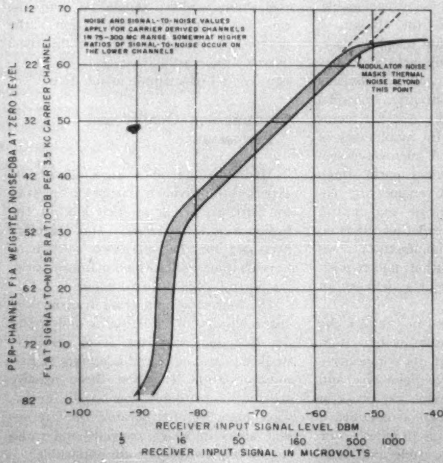


Fig. 4. Thermal noise characteristic of a Lenkurt type-72B 900-mc frequency-modulated microwave receiver

Modulator noise can be minimized by good equipment design, but can never be eliminated. A typical microwave transmitter for telephone use will generate about +14 to +16 dba of such noise.

Intermodulation distortion is a more complex noise source. The linearity of the modulator, demodulator, and wide-band amplifiers under peak-load conditions determines the amount of noise contributed as intermodulation distortion. Loading depends upon the type of carrier system, the peak number of channels in use at one time, and the talker levels.

Complete and accurate information on the performance of the microwave equipment with respect to intermodulation distortion must be provided by the microwave manufacturer. These data are needed by the transmission engineer before he can design a system consistent with the established transmission standards. A typical microwave system loaded with 120 telephone channels, operating in accordance with accepted loading statistics, will contribute intermodulation noise of about +16 dba to each carrier channel during 1% of the busy hour.

Thermal noise in the head end of the receiver is caused by the random motion of electrons in the input resistance and in components of the receiver located before the first amplifier stage. For an adequately designed microwave system, thermal noise will have no significance during normal propagation conditions. Thermal noise has to be taken into account only during the time that fading occurs. The thermal noise data for a receiver should be provided by the manufacturer. A set of curves such as that illustrated in Fig. 4 will give complete information.

When the noise characteristics of the channelizing equipment, the modulator noise, the intermodulation distortion, and the thermal noise characteristics of the

Fig. 5. Array of microwave equipment and its associated remote-control and supervisory equipment

Table II. Total Number of Lenkurt Systems in the 890- to 940-Mc Band

Companies	In Operation	Under Construction	Total
Independents	11	7	18
Bell Telephone System	11	2	13
	22	9	31

Table III. Total Number of Sections in the 890- to 940-Mc Band

Companies	In Operation	Under Construction	Total
Independents	15	22	37
Bell Telephone System	21	3	24
	36	25	61

Table IV. Maximum Number of Tandem Sections in One System

Companies	In Operation	Under Construction
Independents	4	8-tandem section with 5 single-section spurs
Bell System	4	2

Table V. Average Number of Tandem Sections in One System

Bell Telephone System	Independents	All Users
1.85	1.78	1.81

Table VI. Repeaters

Companies	Branching or Drop-Off	Through	Total
Independents	7	7	14
Bell Telephone System	3	9	12
	10	16	26

Table VII. Typical System Statistics

	Section Length, Miles	System Length, Miles	Tower Height, Feet
Maximum	56.0	190.0	440
Minimum	5.5	13.5	
Average	23.0	42.0	106

microwave equipment are known, the noise performance of the over-all system can be calculated for both normal and faded propagation conditions. With noise factors of the magnitudes described, microwave toll systems employing single-side-band suppressed-carrier channelizing can be extended to 10 or 15 repeater sections without resorting to the use of compandors or heterodyne repeaters. Section lengths up to at least 55 miles are possible, depending upon the type of terrain to be traversed and the clearance which can be obtained.

Heterodyne repeaters, while capable of reducing noise in long systems, lack flexibility for dropping and inserting channels at repeater and branching points. Such flexibility has proved advantageous in many medium-route installations. Furthermore, maintenance is standardized when back-to-back microwave terminals are used as repeaters.

Reliability

To provide the high degree of reliability required by the telephone industry, the microwave equipment must utilize the highest quality of components. Adequate metering must be provided, power supplies should be conservatively rated and adequately fused, and the entire system should be provided with monitoring and alarm circuits so that troubles can be localized immediately. Equipment arrangements should be such that all components and panels are easily replaceable.

Where equipment is installed at unattended locations, and particularly where the need for exceptional reliability is important, a suitable remote-control and supervision system should be installed. The remote-control and supervision system should operate over a facility independent of the system it supervises. In the Bell Telephone System, the supervisory and control signals of many *TD-2* routes are carried over paralleling open-wire, cable, or radio systems.

Frequently, medium-route systems do not have adequate paralleling facilities, and may be required to carry their own remote-control and alarm tones. Since a working system is needed to convey supervisory information and, since the most important information concerns unserviceable equipment, stand-by microwave becomes essential. In fact, for maximum reliability of service, stand-by microwave equipment and an adequate remote-control and supervision system go hand in hand. Fig. 5 shows an array of microwave equipment and its associated remote-control and supervisory equipment for use in the telephone network of the Hawaiian Islands.

Reliability is also dependent upon freedom from fading. The results of extensive operating experience, particularly in the vicinity of 900 mc, have made it possible to engineer systems with conservative repeater section lengths up to 55 miles which are almost completely free of fades.

Performance records of 99.80% for single-section systems without stand-by and 99.98% for multisection systems with stand-by are being achieved by operating companies using routine maintenance procedures. Of course, the higher degree of reliability can be obtained only when stand-by microwave equipment is installed, particularly where unattended installations are involved. When these performance records were being determined, the systems were charged with the outage time caused by equipment failure, the traveling time to get to the unserviceable equipment, and usually with the time the systems were turned down for routine servicing.

The requirement for long life of microwave equipment is related to the requirement of reliability. A telephone company should expect its microwave equipment to serve in the toll plant for at least as long as its associated carrier telephone and telegraph equipment. This imposes an obligation upon the microwave manufacturer to design and supply equipment of good quality, and then be ready and willing to stock spare parts and to provide spare panels for an indefinite number of years, even if manufacture of the equipment has been discontinued.

Compatibility

Microwave equipment specifically designed for use in the telephone industry must include features not usually found in radio equipment used for other types of service. These features include:

1. The same general mechanical characteristics as in other telephone equipment.
2. Circuit design and layout in accordance with conventional telephone practices.
3. Electrical characteristics co-ordinated with other telephone equipment.

The physical arrangement of microwave equipment should be such that it can be used anywhere in the toll plant. The advantages of housing microwave equipment without restrictions in any convenient location are immediately apparent. One of the hurdles which radio has had to overcome in the highly standardized telephone industry is its "special" nature. Special equipment, installed on a roof or in a hut, does not take advantage of the constant routine checking and maintenance procedures long established as standard telephone practice.

Experience is proving that microwave equipment, easily accessible in the central office and simply tested, is maintained as easily and performs as reliably as its associated carrier equipment. To make such routine maintenance possible, drawings, engineering information, and maintenance practices along lines standardized by the telephone industry must be made available by the microwave manufacturer.

Conventional telephone engineering practices should be followed in the design and construction of the radio equipment to facilitate installation, investigation of trouble, and maintenance. Also, the equipment should include all necessary safety practices for protecting personnel from the risk of shock.

Electrical co-ordination requires that impedances and levels, particularly on the drop side of the microwave equipment, be identical to those encountered in the connecting carrier equipment. The microwave supplier should provide, on a co-ordinated basis, the necessary junction filters, amplifiers, equalizers, pads, impedance-matching transformers, and other accessory items that permit the telephone company to achieve the utmost flexibility in the use of the microwave equipment. Also, some means should be provided to monitor the microwave circuits, to switch automatically to stand-by microwave equipment, and to include alarms which indicate that the equipment has failed.

The actual cost of maintaining the microwave system must be kept as small as possible, within the standards of reliability. This requires equipment that can be checked and adjusted by telephone craftsmen as part of their routine duties. Maintenance practices should enable nonspecialist craftsmen to localize troubles and make repairs. Here again, it is extremely important that any changes in design be co-ordinated with changes in instruction material so that the equipment and the drawings agree in every detail.

Engineering and Installation Services

In the telephone industry, the engineering load on operating companies is often relieved by placing orders with the supplier for entire communication links. The supplier then takes on the responsibility for engineering the system into the telephone company's plant, ordering all items of equipment, and installing and lining up the system. After completion of line-up, the working system as well as suitable technical data are turned over to the telephone company. At least two suppliers are currently involved in providing such service to the telephone industry.

Statistics on Typical Microwave Systems

It is possible here to provide statistics only on Lenkurt Electric Company, Inc., microwave systems. These statistics should be of some interest however, because they account for more than 40% of the medium-route microwave systems in use by the telephone industry.

Tables I through VII give statistics for the Lenkurt Electric Company, Inc., microwave systems installed or presently under construction. Of 31 systems, 18 have been for independent companies and 13 have been installed by the Bell Telephone System operating companies. The average section length is 23 miles, with the maximum length at 56 miles and the minimum at 5.5 miles. The average system consists of two sections and requires one repeater. Almost 40% of the repeaters have branching or drop-off arrangements, which is a provision not conveniently obtained with heterodyne repeaters. The cost of the microwave equipment may range from about 10 to over 80% of the total cost of the installation. Most systems thus far installed use separate antennas for transmitting and receiving, but several have been equipped with r-f directional filters to permit transmission and reception on a common antenna.

Conclusion

The nature of the telephone industry is such that its suppliers must assume certain immediate and long-term responsibilities that are different from those in other industries. These responsibilities transcend the manufacturing and furnishing of a high-quality piece of radio equipment, and include such factors as design and performance intended specifically for telephone use, the supply of accurate and complete information concerning the system, the guarantee of a continued supply of spare parts and panels, and a continuous up-to-date information service.

Although the initial cost of a microwave system is important, many factors often combine to cause the first cost to lose much of its significance. Such factors are transmission quality, performance, reliability, life expectancy, compatibility with other equipment, and maintenance costs. Regardless of a system's initial cost, unless it fulfills the requirements and continues to do so, it is of little value.

References

1. Load Rating Theory for Multichannel Amplifiers, B. D. Holbrook, J. T. Dixon. *Bell System Technical Journal*, New York, N. Y., Oct. 1939.

2. A Speech Volume Survey on Telephone Message Circuits, V. Subrizi. *Bell Laboratories Record*, New York, N. Y., Aug. 1953.

Traffic Considerations in Community Dial Offices

S. J. LANIGAN
ASSOCIATE MEMBER AIEE

T. J. McDONOUGH
MEMBER AIEE

THE Rural Electrification Administration (REA) under an Act of Congress of October 1949, has the responsibility for making long-term loans for the construction of telephone systems in predominantly rural areas. Security for loans of this type is represented primarily by the ability of the system to provide good service at reasonable rates over the loan period. To insure that the systems for which loans are made represent adequate loan security, it is necessary to establish standards for the design and construction of such systems. It is the purpose of this paper to discuss some traffic engineering standards used by the more than 50 consulting engineering firms who design the systems.

Existing Traffic Objectives

Community dial offices in REA borrowers' systems are usually equipped for conditions anticipated 3 years after cutover. Certain components, battery, charger, building, etc., that are not economically reinforced in a period as short as this will operate initially at lower efficiency so as to last for longer periods. In general, it is preferred that designs be kept convervative because with the planned 3 years' growth there is normally ample time to engineer and install an addition before service deteriorates to the point where there are customer complaints. With the tendency of the suppliers to go more and more to interchangeable components and to jack-in equipment, there is a possibility in multiexchange systems economically to transfer equipment not only within the office but between offices. Intraoffice trunking is based on traffic estimates given in the central office equipment specification. Traffic tables are included in the specification in which the grades of service for the various ranks of switches are as follows:

Linefinders [less than 100% (per cent) lockout]	0.01*
Linefinders (with 100% lockout)	0.02*
Second selectors	0.02
Connectors from second selectors	0.01
Connectors from first selectors	0.03
Revertive call switches	0.01
CLR trunks to local switchboard	0.01

*Approximate.

It may appear to some that these grades of service are better than required to satisfy rural subscribers. It is important to realize that many of the subscribers are on multiparty lines with seven or more others. When a subscriber does succeed in finding the line idle, the probability of finding the called line idle is less than for the average urban system because it too is more than likely a multiparty line. To have the same subscriber receive a busy condition regularly because of insufficient central office equipment will lead to subscriber complaints. Another aspect of this is the fact that even with these criteria, for the average system of REA borrowers, central office equipment (including toll terminal equipment) accounts for only 15% of the total investment in plant.

Since almost 90% of the dial central offices in the borrowers' systems are less than 200 lines, as a general rule the only traffic information furnished the switchboard supplier is the anticipated originating busy hour traffic on a per line basis. In addition to other information about interoffice trunk circuits, the number in the group, their function, and physical length are furnished. The switchboard suppliers compute the interoffice trunk traffic assuming, unless there is information to the contrary, (1) the busy hour condition for the individual ranks of switches occurs at the same time, (2) traffic on 2-way trunks is equal in each direction, and (3) the following table applies:

	Trunk Mileage	
	Less than 8 Miles	More than 8 Miles
Extended area service (EAS)	$P = 0.05$	$P = 0.10$
Toll	$P = 0.03$	$P = 0.05$

Any grading that is required is engineered by the supplier as is the determination of the number of linefinders, selectors (except those associated with the interoffice trunk circuits), and connectors. The reason for leaving this item open is to invite the most economical design of equipment. Some manufacturers furnish 100-point linefinders, others 200-, another effectively 250-, and one has the equipment arranged so that in small offices on outgoing calls the linefinder-connector is released after the trunk has been found and as in the case with other manufacturers, incoming trunks terminated on connectors do not use the local connectors on incoming calls.

Although terminating of trunks on line circuits was very popular among the independents, by 1950 there were several reasons favoring not doing it. The one which had always been present was the fact that incoming trunk calls would then tie up a linefinder as well as a selector and/or connector. The increased usage of these circuits might require an additional intraoffice trunk which would cost more than terminating the trunk on a switch. About 1950, operator toll dialing placed a limitation of one wait for dial tone on incoming calls. This means that where there was a tandem connection to the toll dialing network, either the end office or the tandem office must have its trunks terminated on switches. At about the same time the elimination of manual switchboards caused extended area service to be introduced at a more rapid rate. This, coupled with the introduction of 2-letter, 5-digit numbering and the desirability of universal numbering between dial exchanges, made the use of directing codes and the waiting for dial tones at intermediate points undesirable. The borrower's engineer determines whether or not to terminate trunk circuits on line circuits, selectors, or connectors. There are some installations in REA borrowers' systems where EAS or short-haul toll trunks are terminated on line circuits but usually the distant point has a manual switchboard.

For the past few years all dial switchboards purchased by REA borrowers have had their local first selectors wired for digit absorbing and the majority of them have been equipped for digit absorbing. From now on all local first selectors will be equipped for digit absorbing, although it may not be utilized during the initial period. The borrower's engineer furnishes a numbering plan with the central office equipment specifications and to a certain extent he is able

---

Paper 56-645, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 26, 1956; made available for printing April 17, 1956.

S. J. Lanigan and T. J. McDonough are with the Rural Electrification Administration, Washington, D. C.

to use digit absorbing to reduce the cost of the equipment. An obvious saving is to use it to avoid completely an additional rank of selectors. A more subtle device is to use it to by-pass a rank of selectors. As an example, an incoming toll selector is normally not equipped with digit absorbing. If it happens to be in an office consisting of, say, one or two groups of terminal per line (TPL) connectors and of, say, four or five groups of terminal per station (TPS) connectors, and if the fourth from the last digit of the TPS numbers could be absorbed, the TPS trunk traffic could be routed to connectors off the incoming toll selectors rather than through the local second selectors with a consequent saving in selectors. In some cases the number of levels required for toll, EAS, and local use exceeds ten, the capacity of the switch, so that the engineer is faced with routing either the local traffic or some or all of the EAS traffic through a second rank of selectors. To the extent which is practical, the group with the least traffic should be routed through the extra rank of selectors, because the number of local first selectors will be the same under either plan, but the number of second selectors will depend on the amount of traffic they are to handle.

All connectors (including TPS) are combination toll and local and each is capable of ringing from one to ten (eight for superimposed ringing) parties on the same party line, provided of course all ten parties have been assigned to the same connector group. Normally there will be only one PBX hunting (rotary) group furnished. In large offices lines equipped with lockout may sometimes be segregated at the linefinder end when it is desirable to avoid having all linefinder groups equipped for lockout, and lines may be segregated at the connector when it is determined some should be operated on a TPL basis and the balance of the lines are to be TPS.

Traffic Estimates

In order to determine more accurately the quantity of inter- and intraoffice trunk circuits required, estimates are based on measurements made at such times as to indicate the hourly traffic which may occur with reasonable frequency. The intentions are to provide equipment which will be used occasionally all during the engineered period and efficiently during the busy season just before the next addition. As a practical matter, whether or not this can be realized will depend on local conditions. Considering the difficulty and the cost of gathering data from unattended dial offices, it is recommended that the traffic measurements be taken during the busy season which covers the three or four consecutive weeks when the circuits are busier than at any other similar period. Generally, all of the circuits in the central office will have the same busy season. In the following discussion, this is assumed to be so. If it is not, adjustments for certain trunk groups can be made based on the measured data, in lieu of a switch count at some other time. It is recommended that all traffic registers be read at the same hour of the same day of each of four consecutive weeks during the busy season prior to a conversion or addition. A thorough traffic study should be made during one of these weeks during which switch counts should be taken on the five business days, normally Monday through Friday. Weeks in which there are holidays are avoided. During the switch count week all traffic meters should be read at approximately the same time each day, preferably the first thing each morning. Measurements should be taken during the busiest hour of each of the five business days. One of the several methods of making switch counts, i.e., physical observations, pen recorder, or automatic trunk usage equipment, should be used on each trunk group to be measured. The observation period should be sufficiently long to insure that the busiest hour of the day will be included. In addition to the weekly and daily reading of the traffic registers, they should also be read at the beginning and end of the busy hour. While there is some stability in trunk groups of ten or more trunks, even here it is a mistake to presume that the hour found busiest on the first day will be the busy hour on succeeding days. With groups of two or three trunks errors of more than 100% of the first day's busy hour readings have been observed. The reason for recording the traffic register readings is twofold: first, to highlight trouble conditions which could cause false indications of usage; and second, to provide a means for averaging the 5 days' busy hour data over the three highest weeks of the busy season. Most of the central offices furnished to REA specification have peg-count meters on connector groups only. It is recommended that all traffic in the office be presumed to be proportional to the connector peg count. A check of the reasonableness of this assumption can be made by comparing the busy hour data for the 5 days in which counts were made. Of course, if peg-count meters are furnished for other trunk groups, the results will probably be more reliable if the usage of these trunk groups are presumed proportional to the associated peg-count meter rather than to the connector peg count. Note that the usage registers (ATB, LTB, and OF) are not directly proportional to usage but they serve as a check of the result. The ratio of the busy hour registrations to the week's registrations during the week in which the switch counts were made should be satisfactory for checking the busy hour estimate for the other 3 weeks in the busy season. The results of this study should then be used as a basis for estimating future requirements.

In manual offices and attended dial offices it might very well be as economical to take switch (or cord) counts during the busy hour of 2 or 3 days in each week of the busy season as it was to obtain the foregoing data for the unattended dial office.

In some situations, a manual switch (or cord) count is the most economical method of getting the usage data from an existing office. When a complete study of the office, rather than just one or two trunk groups, is to be made, the manual switch-count method is less satisfactory. In rural communities, the busy hours of interoffice and intraoffice trunk groups often are not the same, which means that switch counts may have to be taken at various times during the day for different trunk groups. In some cases this may be quite satisfactory but usually it is not.

Pen recorders with from 1 to 20 pens each have been used to study traffic for longer periods of time. When the circuit associated with an individual pen becomes busy the pen moves about 1/8 inch. A pattern is traced in ink on a moving paper tape 103 feet long which is calibrated and synchronized. While this instrument can be made to run the tape at various speeds, it seems best to operate it at 1 foot per hour which means that if it is started at 8 a.m. Monday, it will run until Friday afternoon, which for the vast majority of offices will include the busy hours for five business days. This instrument gives an excellent record which makes it possible to discount circuits which it may indicate were in trouble during test. Each recorder can monitor from 1 to 20 circuits in one or more trunk groups. The fact that detailed information is available for the individual trunk circuit makes it especially useful in correcting difficult problems in graded multiple. One objection to its use is that it takes considerable time to analyze the data satisfactorily.

Another instrument similar to those available commercially but designed and built by REA has been found very useful in traffic measurements. It uses a 10-bank 25-point rotary switch and 11 traffic registers. One register is associated with the wiper of each bank of the switch and the 11th indicates the number of observations made during the study period. One observation is made per minute. This instrument is capable of studying from one to ten groups of trunks with from 1 to 25 trunks in a group. It has one of the limitations of the manual switch-count method in that the meters require reading once or twice an hour at about the time of the busy hour, although it operates unattended at other times. To make this instrument even more valuable, a 16-millimeter black-and-white motion picture camera automatically takes a picture of all meters every 30 minutes. With this extra equipment this instrument can also be set up early Monday morning and taken down Friday evening, making it especially useful for unattended dial offices.

If either the pen recorder or the automatic traffic measuring set is operated unattended, it will not be practical to get traffic register readings during the busy hour but it will still be possible to compare traffic during the 4 weeks on the basis of the weekly registrations.

In many instances there are no historical data to help establish the busy season. It is recommended that at least during the first year of dial operation traffic registers be read and recorded each time an unattended office is visited. This should insure a set of readings for each month of the year. It is also recommended that each year during the busy season some time be spent in the analysis of traffic register readings. An annual switch count does not seem justified except when there is a likelihood of an addition or conversion being required.

REA's central office equipment specification requires that the following registers be furnished, and in offices over 600 lines the registers are to be grouped in one location:

Connector peg count

Connector overflow (OF), last trunk busy (LTB), or an all trunks busy with 6-second timing (ATB-6")

Trunk OF, or ATB-6", on all nongraded trunk groups including intraoffice trunks

Trunk OF or LTB on all graded trunk groups including intraoffice trunks

The REA has recently completed tables which are a modification of the widely used truck adjustment tables. They give the anticipated number of registrations for the various trunk groups under the estimated load. One of the shortcomings of the tables is that except for the timed ATB table the average holding time must be known. Except for connectors no other trunk group is equipped with peg-count registers so it is not convenient to estimate holding time. The borrowers are cautioned that the number of registrations in the table are to be used only as an indicator that there may be an overload, or for that matter an underload. Except for the timed ATB registers the number of anticipated registrations is inversely proportional to the average holding time. Since the initial installation is engineered for a 3-year period, it is likely that shortly after cutover the traffic usage meters on small trunk groups will not be registering even during the busy hour. Even though it is recognized that the ratio may change as the trunk group approaches full load, it is recommended that as soon as practical the borrower determined the ratio of the monthly, weekly, and daily registrations to the busy hour. Many of the small dial offices will be visited only two or three times a month so such ratios will be of value. The REA has tried to impress on the borrowers the fact that the absolute number of registrations is not in itself indicative of the service offered. The grade of service desired, the average holding time, and the number of trunks in the group all help determine the number of registrations. When the registers indicate an apparent overload it is recommended that one of the three methods of measuring usage discussed previously be used.

In some cases existing toll traffic is to be converted into EAS traffic. Such traffic is not measurable ordinarily. The recommended way of estimating it is to determine from traffic records the average number of toll messages per busy hour per main station. Adjustments will probably be required if there is to be a sizable growth in the number of subscribers, especially if a few individuals or firms make a large number of calls. The elimination of the toll charge and improved service will cause a stimulation in calling. The amount of the increase depends somewhat on such items as the distance, toll charge, county seat, schools, shopping, and hospitals. To a certain extent the average holding time depends on these factors also. Adjustments similar to these will also be required to estimate toll traffic when extended area service is added or discontinued.

Where existing manual service is not satisfactory due to overloaded party lines, poor transmission, or insufficient central office equipment or personnel, traffic measurement will probably not be worth taking. When experience with similar offices in the area is not available, the following estimates of originating traffic will usually give satisfactory results:

1-party business	1.5 unit calls (CCS)
1-party residence	1.0
2-party business	1.2
2-party residence	0.8
4-party station	0.6
Multiparty station	0.4–0.5
Pay station	1.5
PBX trunks	4.0

Recent Changes

Until recently the vast majority of the interoffice trunk circuits were 2-way, even in groups of 20 to 25. Beyond this there would usually be two groups of 1-way trunks. REA recently began promotion of combinations of 1-way trunks as first choice and 2-way trunks as last choice in groups of from 11 to 35 trunks. Apparently they were not being specified in the past for two major reasons. First, on new installations the prediction as to how the traffic is going to split is a difficult one so that one large group of 2-way trunks is insurance against a bad estimate. Second, there was no specific recommendation as to how to use the tables for combination groups. Combination trunk groups to be engineered as a graded group with one or more legs of the grade in each central office are now being advised. The 2-way trunks are made last choice in both offices and the 1-way trunks become the first choice trunks. As an example of the type of saving being contemplated, consider a situation in which in the past 25 2-way trunk circuits and 25 incoming selectors would have been required. Under the new plan only ten 1-way, and five 2-way trunk repeaters and 15 incoming selectors would be required in each office if loop dialing circuits had been specified. If some other signaling method had been used, ten additional incoming trunk circuits would have been required in each office.

The following grading principles have been emphasized:

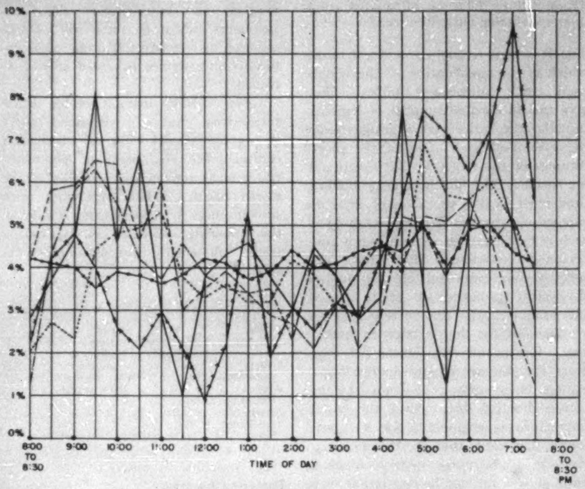


Fig. 1. Traffic on linefinders (hourly variation)

1. The number of trunks common to each leg of the grade should be equal to or greater than one half of the access. Normally there is 10-point access in which case there should be a minimum of five trunks common.

2. As nearly as practical, each leg of the grade should offer the same amount of traffic. This objective can sometimes be met by having more than one shelf of selectors in a leg of the grade.

3. In general, the more legs of the grade (up to about seven or eight) there are, the more efficient the grade. The number of legs in a particular subgroup depends somewhat on the number of shelves and also whether a balanced load is offered by each shelf. Usually an even number of legs in the grade is desirable so that a reversal in the common group in the middle of the grade will better utilize the equipment.

The following formula gives the number of trunks individual to each leg. Fractional trunks in this formula indicate that some trunks must be shared with one or more other legs:

$$\frac{\text{Number of trunks—access}}{\text{Number of legs—1}} = \text{Number of individual trunks}$$

In the case of the 25 trunks referred to previously, assume they are connecting step-by-step offices with approximately equal traffic in each direction (study only even number of legs).

$$\frac{25-10}{4-1}=5 \qquad \frac{25-10}{6-1}=3$$

One leg in each office was not checked because there were more than 15 trunks to be graded. Two legs of the grade in each office is satisfactory. With three legs in each office one additional selector (and possibly trunk circuit) would be required although the group would be more efficient.

Another reason for placing more emphasis on grading at this time is because the number of offices in the REA program of 1,000 lines or more is increasing steadily. Even though the suppliers are presently engineering the grading, an understanding of the technique and the principles is required in order for the engineer satisfactorily to perform acceptance tests. Almost as important as getting an acceptable grading pattern is getting one from which it is easy to grow. It normally would not be desirable to have one that required a major revision when one or two trunks were added.

In some instances, traffic tables which assumed full availability were used in determining the number of trunks in graded groups. To avoid this, the interoffice trunk table was modified so that it is now ungraded up to ten trunks and graded beyond ten. Also, a recommendation that was frequently forgotten is that when there are more than ten trunks in an extended area service group, a grade of service of $P = 0.05$ (rather than $P = 0.10$) should be used even though the distance might be greater than 8 miles. With this in mind, no reference is now made to $P = 0.10$ for groups of 11 or more trunks.

A requirement has recently been added to the central office specification that all intraoffice trunk groups have usage registers installed. There had been instances

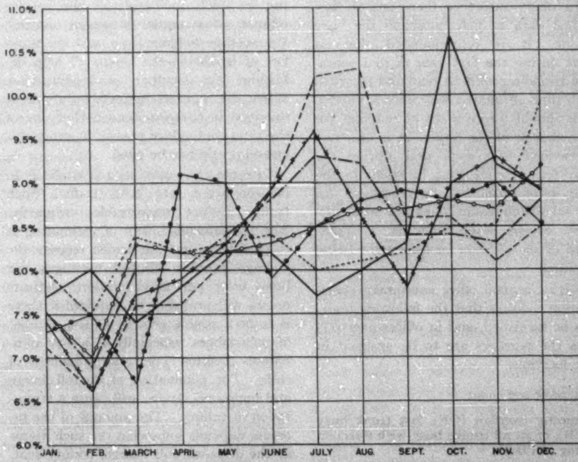


Fig. 2. Traffic on linefinders (monthly variation)

of complaints from subscribers before traffic registers gave an indication of trouble, but this is less likely to happen when intermediate selectors are metered. In one case it was not necessary to reinforce a trunk group when it was found the overload was in a small group of second selectors. Such troubles may be overlooked even by someone trying to do a good traffic study if there are no meters to flag the problem.

It is important that the engineers exercise good judgment, especially in traffic studies. In the design of REA systems, very frequently the individual who makes the commercial survey makes the measurements, prepares the traffic estimate, designs the outside plant, and prepares the central office equipment specification. Such an individual is in a position to prevent optimism or pessimism from being compounded. He can and should make adjustments to give the best over-all design. As an example, some carrier equipment comes in subgroups of four channels each. The fourth channel in the group is relatively inexpensive. The first channel in the next group is quite expensive because of equipment common to four channels. The number of trunks to be installed initially should be determined by such facts rather than by the result of multiplying a certain number of unit calls per station by the number of stations estimated 3 years after cutover. If the engineer knows that there are ample cable pairs, it is possible to give improved service for the same central office cost. In going from, say ten 2-way to seven 2-way, three 1-way incoming, and three 1-way outgoing trunks, there will be a saving (if any change) in central office equipment cost. In many systems the probability of continued growth is great but there are some where growth is not likely. It is important that the designer have this in mind when making borderline decisions.

Results of Traffic Studies on New Installations

The REA has made detail traffic studies in more than 50 central offices. Most of the data compiled so far have been done with the 20 pen recorders. Some examples of the distribution of traffic have been included. Fig. 1 shows how the local originating traffic (unit calls) in six different central offices varies during the day. Fig. 2 shows how the local originating traffic (overflow registrations) in seven different central offices varies each month during the year. Notice that in some offices the busy season extends over several months while in others there is a severe peak.

Based on these studies, the average originating traffic runs from about 0.5 unit calls per station in the small rural offices with little or no EAS traffic to 1.5 unit calls per station in offices with EAS to large metropolitan areas. Of this originating traffic, 20 to 50% will be outgoing EAS and 5 to 25% outgoing toll. The highest EAS and lowest toll rates usually go together and to a lesser degree the reverse is true. It no longer seems strange to have an averaging originating traffic of three unit calls per line, although a few years ago it was almost unheard of in REA designs.

An REA questionnaire requesting that the borrowers tabulate and furnish monthly readings of the traffic registers for the previous 12 months did not have as good a response as was hoped; 23 companies operating 57 exchanges replied. An analysis of the results did not develop facts which would warrant change in basic recommendations. For the most part, those individuals who replied understood their traffic problem and would include footnotes explaining an unusual number of registrations. When overloads were obvious, plans had already been made to provide relief.

Looking Ahead

It is probable that in the next year or two sufficient data can be compiled so that a pattern of regional differences might be determined. Such information may even become more reliable than traffic estimates for dial switchboards based on studies of traffic on manual switchboards. It is not felt that such averages would be suitable for engineering additions to dial offices. Instead actual traffic studies will be required for some time to come. At about the time well-tested traffic estimates can be used in lieu of traffic measurements on manual switchboards, the consulting engineers will probably have become accustomed to and proficient in taking traffic measurements so that the need will be much less than at present.

Some of REA's future efforts will be spent in demonstrating to the borrowers the value of traffic information as a management tool. As more and more of the projects outgrow the initial installation, management becomes increasingly aware of the need for traffic data

The Rural Electrification Administration will continue to evaluate its standards in the light of changing conditions and it is prepared to make adjustments in these standards if it is found they are no longer suitable.

References

1. Basic Theory Underlying Bell System Facilities Capacity Tables, A. L. Gracey. *AIEE Transactions*, vol. 69, pt. I, 1950, pp. 238–44.

2. The Reliability of Telephone Traffic Load Measurements by Switch Counts, W. S. Hayward, Jr. *Bell System Technical Journal*, New York, N. Y., vol. XXXI, Mar. 1952, pp. 357–77.

3. General Aspects of Community Dial Office Equipment, Albert Burkett. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 1540–43.

4. Some Engineering Considerations in the Design of Telephone Systems to Serve Predominantly Rural Areas, T. J. McDonough, Warner T. Smith. *Ibid.*, vol. 73, pt. I, July 1954, pp. 208–14.

◆

Discussion

Imre Molnar (Automatic Electric Company, Chicago, Ill.): This paper contains much helpful information for traffic engineering of small central offices. It is gratifying to learn that the REA and the authors prefer to base engineering on traffic studies rather than average traffic estimates and feel that traffic usage registers prove in their worth because their results are more tangible. But where such studies are not feasible the table of suggested originating traffic estimates contains valuable information. The plea for good judgment on the part of engineers to prevent optimism or pessimism from being compounded is a particularly sound advice. In traffic engineering the temptation is great to follow the book and be contented when the sliderule produces numerical results; the traffic engineer must translate at each step the numerical result into terms of physical equipment and form a mental picture of the events and reasonableness of results. The suggested methods for traffic estimates are practical and should provide sufficient foundation for engineering community dial office trunking. Perhaps the authors are unaware that there are means available to synchronize peg-count meter readings with automatic traffic recorders.

There are a few details of the paper which appear to require further explanation.

1. It would be interesting to know the considerations governing the grade of service specification, which differs considerably from those in general use. Since the cost of the connector and linefinder (with associated first selectors) ranks are very substantially higher than others, they should stand the greatest amount of the loss when apportioning the total grade of service, instead of the second selectors. This is particularly true for linefinders, where a small increase in delay for dial tone is permissible, particularly when subscribers on multiparty lines

are accustomed to service restrictions on account of sharing the line with many other parties. Likewise, the reverting call switches should be much less liberally provided, because it is usually only preceded by the linefinder delay, and also because once the reverting call switch is reached, the call is subjected no further to line busy.

2. The reason given in the paragraph following the grade of service specification is not clear. If the authors' reasoning is that the probability of finding a multiparty line busy is greater than that of finding a single- or 2-party line busy, and that therefore this situation should not be aggravated by busy-in trunk equipment, it seems that with the usual line-busy rate on multiparty lines even a substantial increase of trunking loss would be relatively small in proportion, and the single- and 2-party subscribers would carry an undue proportion of extra cost in the trunking for the benefit of multiparty subscribers.

3. Why is extended area service trunking less liberal than toll trunking (usually the opposite is recommended)? Is it because toll trunks carry pay-as-you-go revenue traffic (in both directions)? However, the extended area trunking is supposedly paid also by increased monthly rates.

4. What is the distinction between connector overflow, etc., and intraoffice trunk overflow? And why are no registers specified for linefinders?

5. A split between 1-way and 2-way trunks is in many applications a very worth-while feature. However, it must be realized that 20 1-way and 5 2-way trunks are not equivalent to 25 2-way trunks, but that there is always a trunk penalty which must be weighed against the saving in central office equipment, and even more so when there is a difference in the respective trunk busy hours.

6. The authors' request for a grading pattern that is easy to grow from is very sound, but it should also be realized that their grading principles 1 and 3 do not lead to the most efficient grading. Both theory and experience have shown that the most efficient grading can be obtained, first, if the total number of switch outlets in a grading is about twice the number of connecting trunks, and second, if the transition from individuals to commons is as smooth as possible—that is, when it is done in approximately equal number of steps. For instance, if 560 unit calls are offered to a 4-legged grading consisting of four steps of individuals, three steps of pairs, and three steps of quadruples, the overflow is theoretically a little over one half of that obtained by the grading suggested in the paper with 25 trunks, although the latter is easier to administer for growth. In what way is a better utilization of equipment expected by a reversal of the common group? The traffic-carrying capacity cannot be increased by such reversal; only the average number of hunting steps and chance for double connection will be somewhat reduced.

S. J. Lanigan and T. J. McDonough: We appreciate Dr. Molnar's comments, and will try to clarify the questions he has raised.

1. We will probably adopt Dr. Molnar's suggestion of providing connectors and reverting call switches on a less liberal basis. The grades of service for the various ranks of switches deviate only slightly from Bell System practice. Bell System experience with linefinders proved to them that a grade of service as poor as $P = 0.04$ was not acceptable to the public. In unattended dial offices of small companies the administration and maintenance will generally not be as good as in the large attended offices. Load unbalance between subgroups, delay in restoring service, and the fact that many subgroups use 100-point switches influenced us in deciding upon a grade of service after dial tone of $P = 0.03$ for local calls.

2. In most REA systems the rural subscribers make up a majority of the total subscribers. The rates paid for other grades of service depend on the rural residence rate. We would say that the rural residence subscriber is getting the grade of service he requires. The 1-, 2-, or 4-party subscriber may be getting somewhat better service than required.

3. We have been told that on local calls the Bell System tries to maintain a grade of service after dial tone of $P = 0.04$. In multioffice exchanges this means giving a better grade of service on interoffice trunks than we advocate for extended area service. In REA borrower's systems, trunk groups are generally small, and the recommended grades of service are satisfactory. In instances we know of where subscribers have complained, the grades of service have been worse than $P = 0.10$. In addition to the fact that toll traffic is pay-as-you-go, toll trunks handle emergency traffic and the introduction of direct distance dialing also justifies the better grades of service on toll groups.

4. We neglected to state that an OF, ATL, or ATB-6" timing meter is required for each linefinder group. In connection with traffic meters, "intraoffice trunks" refers to second and third selectors and revertive call switches.

5. No comment.

6. We are soon to issue information on engineering central offices larger than 800 lines which will include grading patterns standard in the Bell System for step-by-step offices. The transition from individual to common trunks occurs on the first five contacts, and is not always smooth; however, we plan to follow the Bell System in trying to maintain a minimum of five common trunks in the grade because we expect that a high percentage of graded subgroups will have some unbalance.

Vol.
Pt. 1
195

Dynamax, a New Crystal and Domain-Oriented Magnetic Core Material

G. H. HOWE
NONMEMBER AIEE

Synopsis: Dynamax, a new high-permeability magnetic material containing approximately 65 per cent (%) nickel, 2% molybdenum, and 33% iron, has been developed in the form of thin tape for toroidal cores. It has dynamic characteristics which are superior to those of any rectangular hysteresis loop materials commercially available today. The most promising potential applications are in high-performance magnetic amplifiers and possibly in special transformers. In addition, its d-c magnetization characteristics are comparable to those of Supermalloy except that the Dynamax has a rectangular hysteresis loop and higher remanance. Dynamax has d-c properties which are superior to Deltamax except for lower saturation.

This development is the result of an alloy modification of 65 permalloy to raise the specific resistance to a useful range and a considerable amount of processing research to produce controlled textures and domain patterns by rolling from specially cast slab ingots, primary recrystallization, pure dry hydrogen annealing, and heat treatment in a magnetic field.

SEVERAL important developments in the field of magnetic materials have come from the nickel-iron system, notably the permalloys. Outstanding from the point of scientific interest were the results reported by Bozorth, Dillinger, and Kelsall[1] on their investigation of the effect of heat-treating a series of nickel-iron alloys in a magnetic field. A large increase in permeability developed, the maximum value occurring in an alloy of about 65% nickel where the Curie temperature is highest in this nickel-iron composition range. Dahl and Pawlek[2] confirmed the large effect of cooling in a magnetic field using 12-mil-thick nickel-iron strips both with random and preferred crystal orientations. Subsequent development of the alloys of this type and the controversy over the mechanism of magnetic annealing have been discussed by Hoselitz.[3]

Fig. 1 (left). Cross section of vacuum-cast Dynamax ingot

Fig. 2 (right). Cross section of $1^1/_2$- by 4-inch Dynamax slab cast in graphite mold

Experimental

The initial work in this investigation was directed at improving the magnetic properties of 65 permalloy by making grain-oriented thin strip for test in wound toroids.

The alloy was vacuum melted in an induction furnace and poured into a cast iron mold. To improve rolling, 0.3% manganese was added. The ingots were reduced to thin strip by hot forging followed by hot and cold rolling.

After slitting to the desired width, the strip was coated with a thin film of hydrated magnesia and wound into toroidal test specimens. Following a 1,100 C (degrees centigrade) preliminary anneal in pure dry hydrogen, the specimens were heated to 650 C in ordinary hydrogen and cooled in a magnetizing field of 10–12 oersteds at 150 C per hour. This treatment is called a magnetic anneal in this paper. It should be noted, however, that it is slow cooling through the Curie temperature in a magnetic field.

The d-c magnetic properties shown in Table I are typical of those obtained on $1^1/_2$-inch inside diameter by $1^7/_8$-inch outside diameter cores wound with 1/2-inch-wide tape of different thickness and heat treated as discussed. It is observed that:

1. Of the 5-mil specimens, nos. *1*, *4*, and *5*, maximum permeability improved with increased percentage of cold reduction.

2. Referring to the second group of specimens, nos. *3*, *5*, and *6*, for a constant amount of cold rolling, $87^1/_2\%$ cold reduction, the magnetic properties were improved by a decrease from 10 mils to 3 mils in finish thickness. The results of X-ray diffraction tests furnished evidence that the degree of preferred crystal orientation also increased with thinner material.

3. The data on specimen *6B* illustrate the effectiveness of the magnetic anneal on this alloy. The maximum permeability is raised by a factor of 125 and the d-c hysteresis loop is squared.

The extreme degree of magnetic softness attained in 65 permalloy was not useful in a-c applications owing to the low (25-michrohm-centimeter) resistivity of the binary alloy. To remedy this serious limitation, molybdenum was first added to 65 permalloy to increase its resistivity as shown in Table II.

While increased resistivity was attained at some sacrifice in d-c magnetic properties, reasonably high maximum permeability and low coercive force were measured on 2-mil by 1/2-inch-wide tape from *VH 466* containing 2.16% molybdenum, i. e.: specimen *466-4*—338,000 maximum permeability, 11,740 B_r, 0.029 H_c, 47.7 microhm-centimeter. The dynamic hysteresis loop of this specimen at 60 and 400 cycles was found to be narrower than that of Deltamax and 65 permalloy. The 3% material was no better dynamically and had a lower saturation density.

In addition to 60- and 400-cycle dynamic hysteresis loops, specimen *466-4* was tested in a half-wave self-saturated magnetic amplifier circuit and control characteristic curves made at 60 and 400 cycles. The results showed a marked superiority in relation to comparable data on Deltamax cores except that the available power output range of the Dynamax core was only 75% that of the Deltamax core. This difference is attributed to the lower saturation density of Dynamax.

Molybdenum was chosen as the alloying material for the following reasons:

1. It was known to raise resistivity more effectively than silicon for a given reduction in saturation density.

2. Its addition to 79 permalloy had produced Supermalloy whose dynamic properties were excellent.

3. It might provide a means for easy nucleation of the domains and thereby improve the dynamic properties.

Metallurgical Development

The development of good magnetic quality in Dynamax parallels that of producing other grain-oriented alloys in that variations in melting, casting, and hot working are reflected in the quality of finished strip.

Early in this investigation ingots of square cross section were reduced to billets by hot forging and further to strip by hot and cold rolling. Good magnetic properties could not be reproduced consistently, so attention was focused on the structure of the ingots. A typical structure of the cross section of these ingots is shown in Fig. 1.

Table I. D-C Magnetic Properties of 65 Permalloy

Heat	Specimen No.	Thickness, Mils	% Cold Reduction	Normal Induction		Hysteresis, Peak H = 12			X-Ray Diffraction
				Maximum Permeability	At B	Peak B	B_r	H_c	
VH 506	1	5	98	976,000	7,750	13,375	12,700	0.0103	
	4	5	94	782,000	9,500	13,430	12,720	0.0111	
	5	5	87.5	700,000	10,100	14,080	12,400	0.0147	
VH 506	3	10	87.5	555,000	10,000	13,480	12,160	0.0143	No cube texture to any degree
	5	5	87.5	700,000	10,100	14,080	12,400	0.0147	Some cube texture
	6	3	87.5	1,027,000	9,400	13,350	12,840	0.0115	Primarily cube texture
VH 506	6B	3	87.5	8,420*	5,400	13,370	1,200	0.171	
				1,016,000†	10,100	13,530	13,020	0.010	

*4-hour 1,100 C pure dry hydrogen preliminary anneal only.

†After magnetic anneal.

Paper 56-681, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted September 7, 1955; made available for printing April 27, 1956.

G. H. Howe is with the General Electric Company, Schenectady, N. Y.

Grateful acknowledgment is made to W. R. Hibbard, Jr., for his support and advice in this work and to H. W. Lord for his invaluable assistance in evaluating the dynamic properties of the magnetic toroidal core specimens.

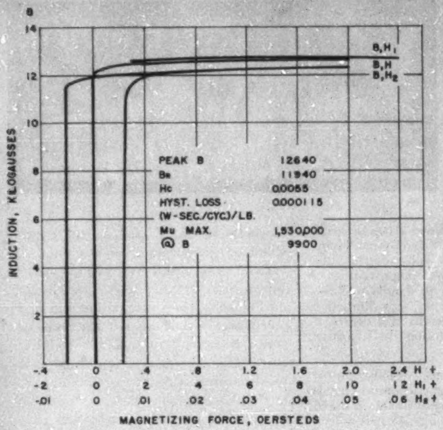


Fig. 3. D-c hysteresis loop Dynamax 2-mil by 1/2-inch tape specimen no. 3B3

It was conceivable that an important variable lay in converting this columnar-cast structure into a uniformly optimum-wrought structure through progressive working under the locally applied forces of a forging hammer. Therefore, a $1^{1}/_{2}$- by 4- by 12-inch slab was cast in a split graphite mold for reduction solely by rolling.

Fig. 2 shows a typical cross section. The ingots being cast vertically, the columnar grains grew from all four sides with most of them perpendicular to the rolling plane.

Table II. Effect of Adding Molybdenum to 65 Permalloy

Material	Molybdenum by Analysis, %	Specimen Resistivity	Specimen Gr.	Curie Temperature, C
VH 465	1	34	8.49	
VH 466	2.16	47.7	8.52	560
VH 561	2.9	54.1	8.63	540

The d-c magnetic properties shown in Table III were obtained on $1^{1}/_{2}$-inch inside diameter by $1^{7}/_{8}$-inch outside diameter cores wound with 2-mil by 1/2-inch-wide tape produced by rolling a portion of the completely columnar slab ingot.

In a previous experiment it was learned that a primary recrystallization treatment improved the magnetic quality. Table III, column 1, shows the effect of this treatment at various temperatures. It is observed that highest permeabilities were measured on tape which had been subjected to recrystallization at 700 C prior to the 4-hour high-temperature anneal and the magnetic anneal.

Reannealing for 4 hours at 1,200 C plus magnetic anneal substantially increased the permeabilities of the prior recrystallized specimens, particularly no. *3B2* which attained 1,615,000 maximum permeability. The poorer properties of no. *3B1* and no. *3B2* after reannealing at 1,300 C can be explained by a difference in crystalline texture.

Table III. Effect of Annealing Treatment on D-C Magnetic Properties of Dynamax

No.	Initial Recrystallization	Maximum Permeability	at B	Peak B	B_r	H_c
	4 Hours at 1,100 C Pure Dry Hydrogen + Magnetic Anneal					
3B1	10 min 500 C	800,000	9,200	12,440	11,490	0.0125
2	10 min 700 C	1,073,000	10,400	12,490	11,910	0.0096
3	10 min 900 C	1,036,000	10,200	12,590	11,980	0.0095
4	None	621,000	8,800	12,740	11,780	0.0150
	4 Hours 1,200 C Reanneal Pure Dry Hydrogen + Magnetic Anneal					
3B1	10 min 500 C	1,231,000	9,300	12,500	11,600	0.0077
2	10 min 700 C	1,615,000	10,100	12,640	12,020	0.0055
3	10 min 900 C	1,530,000	9,900	12,640	11,940	0.0055
4	None	732,000	9,500	12,630	11,880	0.0109
	4 Hours 1,300 C Reanneal Pure Dry Hydrogen + Magnetic Anneal					
3*B*1	10 min 500 C	449,000	8,900	11,090	10,280	0.0176
2	10 min 700 C	551,000	9,500	11,190	10,600	0.0124

Specimen *3B3* measured 1,530,000 maximum permeability after the 4-hour 1,200 C pure dry hydrogen plus magnetic anneals. The d-c hysteresis loop of this specimen is shown in Fig. 3.

The best values which have been attained are 1,780,000 maximum permeability, 11,950 residual flux density, and 0.0053 coercive force. This maximum permeability is higher than any published values the author has seen.

Dynamic Properties

The improvements in dynamic properties obtained by the addition of 2% molybdenum to 65 permalloy cannot be accounted for by the increased resistivity alone. This leads to the belief that the addition of molybdenum brings about a condition in the material which is more favorable to domain nucleation. Because it is believed that imperfections in the magnetic structure of a material are likely spots to form regions of reversed magnetization (i. e., reversal nuclei) it is not unreasonable to assume that the addition of molybdenum may affect the number and distribution of such reversal nuclei.

However, the static and dynamic hysteresis loops recorded by the U. S. Signal Corps on two specimens of Dynamax which had been processed differently indicate that there are processing effects which influence domain size and therefore affect the dynamic loop width. The specimen having the superior d-c properties of low H_c and high ratio of B_r to B_s was inferior dynamically to the specimen having the poor d-c properties. It is apparent that as of now there is no good explanation for this. The answer is expected from further studies of this and similar alloys in search for rectangular core materials having improved dynamic properties.

Dynamax promises to be less sensitive to the effects of temperature variations than are the nickel-iron alloys now commercially available. The effect of core temperature up to 315 C upon the gain of a high-performance magnetic amplifier is about the same for Dynamax as it is for grain-oriented silicon strip cores.

References

1. Theory of the Heat Treatment of Magnetic Materials, R. M. Bozorth, J. F. Dillinger, G. A. Kelsall. *Physical Review*, New York, N. Y., vol. 45, 1934, pp. 742–43.

2. The Effect of Fibrous Structure and Cooling in a Magnetic Field on the Magnetization Curve, O. Dahl, F. Pawlek. *Zeitschrift für Physik*, Berlin-Charlottenburg, Germany, vol. 94, 1935, pp. 504–22.

3. Ferromagnetic Properties of Metals and Alloys, H. Hoselitz. Oxford University Press, London, England, 1952, pp. 232–35.

Discussion

Martin F. Littmann (Research Laboratories, Armco Steel Corporation, Middletown, Ohio): The remarkable properties of Dynamax are largely due to the response of this alloy to the magnetic anneal. By contrast, the regular 48% nickel-iron alloys do not respond readily to a magnetic anneal but values of d-c coercive force of 0.02 to 0.03 oersted together with a B_r/B_m ratio above 0.90 at $H_m = 10$ oersteds are obtainable after annealing at 1,100 C without magnetic annealing.

We note with interest the statement that "the improvements in dynamic properties obtained by the addition of 2% molybdenum cannot be accounted for by the increased resistivity alone." However, no dynamic data are shown for the specimens described in the paper. It would be instructive to compare d-c and dynamic data for specimens of varying d-c properties both with and without the molybdenum addition.

The observation that the Dynamax specimen with a high ratio of B_r to B_s was inferior dynamically to the specimen with poor d-c properties confirms the findings of Ward and Littmann[1] that improved squareness of the hysteresis loop increases the width of the dynamic loop even for the same d-c coercive force.

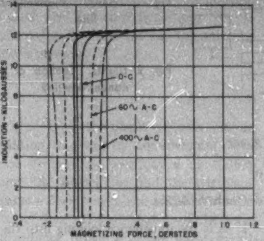


Fig. 4. Hysteresis loops of 2-mil 65 permalloy.

Reference

1. Relation of D-C Magnetic Properties of Oriented 48-Per-Cent Nickel-Iron to Magnetic-Amplifier Performance, C. E. Ward, M. F. Littmann. *AIEE Transactions*, vol. 74, pt. I, Sept. 1955, pp. 422–27.

G. H. Howe: I wish to thank Mr. Littmann for his comments.

Communication with Mr. Littmann has revealed he does not wish to imply that the unusually low values of d-c coercive force, i.e. 0.02 to 0.03 oersted for 48% nickel-iron alloys, are obtainable on all heats. Only certain heats had responded so as to produce this extremely low coercive force and still maintain good rectangularity.

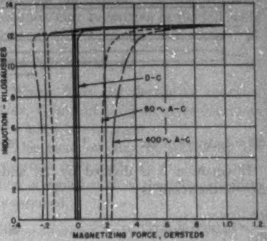


Fig. 5. Hysteresis loops of 2-mil Dynamax

An example of the superiority of Dynamax containing 2% molybdenum over 65 permalloy without the molybdenum addition is furnished in Figs. 4 and 5 which show hysteresis loops for 65 permalloy and Dynamax, respectively. By comparison of the dynamic loops it is observed that the coercive force of the Dynamax is about the same at 400 cycles as that of 65 permalloy at 60 cycles and the Dynamax first quadrant is more rectangular.

Computation of the Impedances of Nonuniform Lines by a Direct Method

LOUIS A. PIPES
MEMBER AIEE

Synopsis: The impedances of nonuniform or tapered lines are computed directly by solving a first-order nonlinear differential equation of the Riccati type. Special cases which include the uniform line, the exponential line, the Bessel line, and the Heaviside-Bessel line are discussed. A general reversion method for solving the nonlinear differential equation for the line impedance is discussed.

TRANSMISSION LINES whose impedance per unit length $Z_L(x)$ and admittance per unit length $Y_L(x)$ vary continuously with the distance x along the line are frequently used in communication engineering. These lines are used to couple circuits that have different impedances in order to avoid abrupt discontinuities. Such nonuniform lines are used in duplex working submarine cables and as circuit elements that have transformer properties.

Several general mathematical analyses of lines of this type have been published.[1–7] The starting points of these analyses are the differential equations for the propagation of currents and potentials of these lines. The solutions are obtained in series form or are expressed in terms of Bessel functions. In certain classes of problems that arise in practice the information concerning the complete distribution of current and potential along the nonuniform line is not required but only a knowledge of the impedance of the line is needed. It is the purpose of this paper to show the line impedance may be obtained by integrating a nonlinear differential equation of the first order of the Riccati type.

The Differential Equation for the Line Impedance

Consider the nonuniform line shown in Fig. 1. Let $Z_L(x)$ be the impedance and $Y_L(x)$ the admittance per unit length of the line. Let Z_r be a lumped impedance connected across the terminals of the line at $x=0$. Let a section of the line of length Δx be represented by Fig. 2.

Let $Z+\Delta Z$ be the impedance across the points A and B of the elementary section of the line of Fig. 2. Z is the impedance across the line at a point x from the receiving end of the line at $x=0$. By the use of the ordinary laws of combination of impedances in series and parallel the following value is obtained for the impedance between the points A and B:

$$Z_{AB} = Z + \Delta Z = Z_L\Delta x + \cfrac{1}{Y_L\Delta x + \cfrac{1}{Z}} = Z_L\Delta x + \frac{Z}{1+Y_LZ\Delta x} \qquad (1)$$

If quantities of order Δx^2 and higher are neglected, the following approximation may be made:

$$\frac{1}{1+Y_LZ\Delta x} = 1 - Y_LZ\Delta x \qquad (2)$$

Therefore equation 1 may be written in the form

Paper 56-685, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted January 19, 1956; made available for printing April 27, 1956.

Louis A. Pipes is with the University of California, Los Angeles, Calif, and the U. S. Naval Ordnance Test Station, China Lake, Calif.

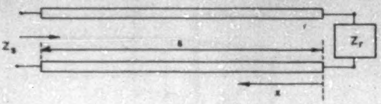


Fig. 1 (left). Nonuniform line

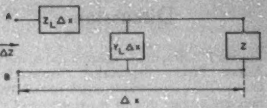


Fig. 2 (right). Section of a nonuniform line

$$\Delta Z = Z_L \Delta x - Y_L Z^2 \Delta x \tag{3}$$

If equation 3 is divided by Δx and $\Delta x \rightarrow 0$, the result is

$$\frac{dZ}{dx} = Z_L(x) - Y_L(x) Z^2 \tag{4}$$

Equation 4 is a first-order nonlinear differential equation of the Riccati type.[8,9] This equation may be transformed into a linear equation of the second order by the following transformation:

$$Z = \frac{1}{Y_L u} \frac{du}{dx} \tag{5}$$

If this transformation is applied to equation 4, the following linear equation in u is obtained:

$$\frac{d^2u}{dx^2} - (Y_L'/Y_L)\frac{du}{dx} - (Z_L Y_L) u = 0 \tag{6}$$

The solutions of the differential equations 4 or 6 will now be obtained for special cases of the variation of $Z_L(x)$ and $Y_L(x)$.

The Uniform Line

If the line is uniform, the impedance $Z_L(x)$ and the admittance $Y_L(x)$ are both constant and have the form

$$Z_L(x) = Z_0 \tag{7}$$

$$Y_L(x) = Y_0 \tag{8}$$

where Z_0 and Y_0 are the constant impedance and admittance per length of the uniform line. In this case, equation 4 may be written in the form

$$\frac{dZ}{(Z_0 - Y_0 Z^2)} = dx \tag{9}$$

and the variables are separated. The integral of equation 9 may be written in the form,

$$Z(x) = Z_c \tanh(ax + \phi) \tag{10}$$

where Z_c is the characteristic impedance of the uniform line and a is the propagation constant of the line. These quantities are defined by the equations

$$Z_c = (Z_0/Y_0)^{1/2} \tag{11}$$

$$a = (Z_0 Y_0)^{1/2} \tag{12}$$

where ϕ is an arbitrary constant. If the line is terminated as shown in Fig. 1, $Z = Z_r$ is at $x = 0$ so that equation 10 gives the result

$$Z_r = Z_c \tanh(\phi) \tag{13}$$

or

$$\phi = \tanh^{-1}(Z_r/Z_c) \tag{14}$$

If Z_s is the impedance looking into the line at $x = s$, then, as a consequence of equation 10,

$$Z_s = Z_c \tanh(\theta + \phi) = [Z_c \tanh(\theta) + Z_r] / [1 + (Z_r/Z_c)\tanh(\theta)] \tag{15}$$

where $\theta =$ as is the angle of the line. This is the well-known result for the impedance of a uniform line of length s terminated by an impedance Z_r. The fact that the impedance of a very long line is Z_c may be seen from equation 10 by allowing x to take a very large value. If this is done, then,

$$\lim_{x\to\infty} Z = \lim_{x\to\infty} Z_c \tanh(ax + \phi) = Z_c \tag{16}$$

Nonuniform Line with $Z_L = Z_0 f(x)$, $Y_L = Y_0 f(x)$

Next to the uniform line, the most simple case to analyze is the one in which the variation of the impedance and admittance per length have the following form,

$$Z_L(x) = Z_0 f(x) \tag{17}$$

$$Y_L(x) = Y_0 f(x) \tag{18}$$

where Z_0 and Y_0 are constants and $f(x)$ is a single-valued continuous function of x. In this case equation 4 may be written in the form,

$$\frac{dZ}{dx} = (Z_0 - Y_0 Z^2) f(x) \tag{19}$$

Hence,

$$\frac{dZ}{(Z_0 - Y_0 Z^2)} = f(x)\,dx \tag{20}$$

If both members of equation 20 are integrated, the result may be written in the form,

$$\frac{1}{a}\tanh^{-1}(Z/Z_c) = \int f(x)\,dx + C \tag{21}$$

where $Z_c = (Z_0/Y_0)^{1/2}$, and C is a constant.

Equation 21 may be written in the alternative form,

$$Z = Z_c \tanh\left[a \int f(x)\,dx + \phi\right] \tag{22}$$

where ϕ is an arbitrary constant to be determined by the termination of the line at $x = 0$. If $f(x) = 1$, the line is uniform and equation 22 reduces to equation 10. Equation 22 may be used to compute the impedance of nonuniform lines for special functions $f(x)$. If, for example, the nonuniform line has a periodic variation of impedance and admittance per unit length of the form,

$$Z_L(x) = Z_0[1 + m\cos(wx)],\ m < 1 \tag{23}$$

$$Y_L(x) = Y_0[1 + m\cos(wx)]$$

then $f(x) = [1 + m\cos(wx)]$ and equation 22 becomes

$$Z = Z_c \tanh\left[x + \frac{m}{w}\sin(wx) + \phi\right] \tag{24}$$

The Exponential Line

A nonuniform line whose performance has been considerably investigated[5,6] is one whose impedance and admittance variations are of the form,

$$Z_L(x) = Z_0 e^{kx} \tag{25}$$

$$Y_L(x) = Y_0 e^{-kx} \tag{26}$$

where Z_0, Y_0, and k are constants. This nonuniform line is called the "exponential line" in the literature. In order to obtain an expression for the impedance of this line, substitute equations 25 and 26 into equation 6 and obtain the following equation,

$$\frac{d^2u}{dx^2} + \frac{k\,du}{dx} - a^2 u = 0,\ a = (Z_0 Y_0)^{1/2} \tag{27}$$

This equation has the solution

$$u = A e^{-kx/2}\cosh(wx + \phi) \tag{28}$$

where

$$w = \left(a^2 + \frac{k^2}{4}\right)^{1/2} \tag{29}$$

and A and ϕ are arbitrary constants. The impedance of the line is given by equation 5 in the form,

$$Z = (1/Y_L u)\frac{du}{dx} = \frac{e^{kx}}{Y_0}[w\tanh(wx + \phi) - k/2] \tag{30}$$

ϕ is an artibrary constant. If the line is terminated at $x = 0$ by an impedance Z_r, then

$$Z_r = \frac{1}{Y_0}[w\tanh(\phi) - k/2] \tag{31}$$

Hence ϕ is determined by the following equation,

$$\tanh(\phi) = \frac{1}{w}[Z_r Y_0 + k/2] \qquad (32)$$

The magnitude of the sending-end impedance of the line Z_s is obtained by placing $x = s$ in equation 30. The result is

$$Z_s = \frac{e^{ks}}{Y_0}[w\tanh(ws+\phi) - k/2] \qquad (33)$$

where ϕ is determined by equation 32. If $k=0$ it can be shown that equation 33 reduces to the case of the uniform line equation 15.

The Case $Y_L = Y_0$, $Z_L = Z_0x^p$

Another special case that admits a relatively simple solution is the one in which the admittance per length Y_L is a constant Y_0 and the impedance per length Z_L varies as a power of x in the form, $Z = Z_0x^p$. In this case, equation 4 takes the form,

$$\frac{dZ}{dx} = Z_0x^p - Y_0Z^2 \qquad (34)$$

Equation 6 is now of the form,

$$\frac{d^2u}{dx^2} - Z_0Y_0x^pu = 0 \qquad (35)$$

The general solution of equation 35 may be written in the form,

$$u = x^{1/2}[AI_n(w) + BI_{-n}(w)] = x^{1/2}F(x) \qquad (36)$$

where

$$w = rx^q$$

$$r = \frac{2(Z_0Y_0)^{1/2}}{(p+2)}$$

$$n = 1/(p+2)$$

$$q = (1+p/2)$$

$I_n(w)$ is the modified Bessel function for the first kind of order n.[10] The impedance Z is given by equation 5 in the following form,

$$Z = (1/Y_Lu)\frac{du}{dx} = \frac{1}{Y_0}\left[1/2x + (1/F)\frac{dF}{dx}\right] \qquad (37)$$

$$= (1/2Y_0x)(1 + 2rqx^q[I_n'(w) + CI_{-n}'(w)]/[I_n(w) + CI_{-n}(w)])$$

where

$$I_n'(w) = \frac{dI_n}{dw}(w) \qquad (38)$$

The arbitrary constant C may be determined from the boundary condition of the line at $x=0$.

The Case $Y_L = Y_0x^a$, $Z_L = Z_0x^b$

The case in which the admittance and impedance per length have the form $Y = Y_0x^a$ and $Z = Z_0x^b$ has been investigated by several writers.[2,4] In general, if $a=b$ the line impedance may be obtained by the nonuniform line method and is given by

$$Z(x) = (Z_0/Y_0)^{1/2}\tanh[x^{a+1}/(a+1)+\phi], \quad a \neq -1 \qquad (39)$$

If $a=b=0$, the line is uniform and the solution is given by equation 10. The case $a=-b$ has been called by Ballantine[2] the Bessel line and the particular case $a=1=-b$ is known as the Heaviside-Bessel line. If $b=1$ and $a=0$ the series impedance tapers and the shunt admittance is constant. The reverse case has also been investigated.

In the general case $Y_L = Y_0x^a$, $Z_L = Z_0x^b$ equation 6 becomes

$$\frac{d^2u}{dx^2} - \frac{a}{x}\frac{du}{dx} - k^2x^{(a+b)} = 0, \quad k = (Z_0Y_0)^{1/2} \qquad (40)$$

The solution of equation 40 may be expressed in the form,

$$u = x^p[AI_n(w) + BK_n(w)], \quad p = (1+a)/2, \quad w = \theta x^q,$$

$$\theta = 2k/(a+b+2), \quad q = 1 + (a+b)/2, \quad n = (1+a)/(a+b+2) \qquad (41)$$

A and B are arbitrary constants and the functions $I_n(w)$ and $K_n(w)$ are modified Bessel functions of the first and second kinds of order n.[11] The line impedance $Z(x)$ may be obtained by means of equation 5 in the form,

$$Z(x) = 1/(Y_Lu)\frac{du}{dx} \qquad (42)$$

where u is given by equation 41 and

$$\frac{du}{dx} = px^{p-1}[AI_n(w) + BK_n(w)] + x^p[AI_n'(w) + BK_n'(w)]\frac{dw}{dx} \qquad (43)$$

If equations 41 and 43 are substituted into equation 42 and certain algebraic reductions are performed, the impedance $Z(x)$ may be expressed in the form,

$$Z(x) = \frac{1}{Y_0}x^a(p/x + \theta qx(p+q-1) [I_n'(w) + CK_n'(w)]/[I_n(w) + CK_n(w)]) \qquad (44)$$

where C is an arbitrary constant which may be determined by the given termination of the line at $x=0$. The primes indicate differentiation of the Bessel functions with respect to their argument w.

Determination of the Line Impedance by the Method of Reversion

In certain cases, the line impedance $Z(x)$ may be obtained approximately by obtaining an approximate solution of the nonlinear differential equation 4. This equation may be written in the following form,

$$DZ + Y_L(x)Z^2 = kZ_L(x) \qquad (45)$$

Where D is the derivative operator d/dx, $Y_L(x)$ and $Z_L(x)$ are given functions of x and k is a constant.

This equation may be solved by the reversion method.[13] To do this, a solution of equation 45 of the following form is assumed,

$$Z(x) = Z_1(x)k + Z_2(x)k^2 + Z_3(x)k^3 + Z_4(x)k^4 + \ldots \qquad (46)$$

The functions $Z_i(x)$ are impedance functions that may be determined by substituting equation 46 into 45 and equating like powers of k in the resulting equation. If this is done, it is found that the $Z_i(x)$ functions satisfy the following differential equations:

$$DZ_1 = Z_L(x) \qquad (47)$$

$$DZ_2 = -Y_LZ_1^2 = f_2(x) \qquad (48)$$

$$DZ_3 = -2Y_LZ_1Z_2 = f_3(x) \qquad (49)$$

$$DZ_4 = Y_L(Z_2^2 + 2Z_1Z_3) = f_4(x) \qquad (50)$$

$$DZ_5 = -2Y_L(Z_1Z_4 + Z_2Z_3) = f_5(x) \qquad (51)$$

$$DZ_6 = -Y_L(Z_3^2 + 2Z_1Z_5 + 2Z_2Z_4) = f_6(x) \qquad (52)$$

$$DZ_7 = 2Y_L(Z_1Z_6 + Z_2Z_5 + Z_3Z_4) = f_7(x) \qquad (53)$$

The equations satisfied by the functions $Z_i(x)$ of higher indexes may be obtained in a similar manner. If the boundary condition $Z = Z_r$ at $x=0$ is used, it can be seen that the integrals of the foregoing differential equations may be expressed in the form,

$$Z_1(x) = Z_r + \int_0^x Z_L(x)dx \qquad (54)$$

$$Z_i(x) = \int_0^x f_i(x)dx, \quad i = 2,3,4,5,\ldots \qquad (55)$$

If the constant $k=1$, it is seen that equation 46 is a formal solution of equation 4 provided the series converges. For example, consider a uniform line so that $Z_L = Z_0$ and $Y_L = Y_0$. Let the line be short-circuited at $x=0$ so that $Z_r = 0$.

In this case,

$$Z_1 = Z_0x \qquad (56)$$

$$Z_2 = -Y_0\int_0^x Z_0^2x^2dx = -Z_0^2Y_0\frac{x^3}{3} \qquad (57)$$

$$Z_3 = 2\int_0^x Y_0Z_0^3Y_0\frac{x^4}{3}dx = \frac{2}{15}Z_0^3Y_0^2x^5 \qquad (58)$$

With $k=1$, the series equations 46 in this case become

$$Z(x) = Z_0x - Z_0^2Y_0\frac{x^3}{3} + \frac{2}{15}Z_0^3Y_0^2x^5 - \ldots \qquad (59)$$

This series may be written in the following form,

$$Z(x)=(Z_0/Y_0)^{1/2}\left[(Z_0Y_0)^{1/2}x-(Z_0Y_0)^{3/2}\frac{x^3}{3}+(Z_0Y_0)^{5/2}\frac{2x^5}{15}-\ldots\right]$$
$$=Z_c\left[(ax)-\frac{(ax)^3}{3}+(ax)^5\frac{0.2}{15}-\ldots\right]$$
$$=Z_c\tanh(ax),\ Z_c=(Z_0/Y_0),^{1/2}a=(Z_0Y_0)^{1/2} \qquad (60)$$

The reversion method is therefore seen to give the series expansion of the impedance of the uniform line. The method is particularly useful if $Z_L(x)$ and $Y_L(x)$ are polynomials in x.

References

1. PROPAGATION OF PERIODIC CURRENTS OVER NONUNIFORM LINES, **J. R. Carson.** *The Electrician*, London, England, vol. 86, no. 10, Mar. 1921, p. 272.

2. NONUNIFORM LUMPED ELECTRIC LINES, **S. Ballantine.** *Journal*, Franklin Institute, Philadelphia, Pa., vol. 203, 1927, p. 561.

3. SINUSOIDAL CURRENTS IN LINEARLY TAPERED LOADED TRANSMISSION LINES, **J. W. Arnold, P. F. Bechberger.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 19, 1931.

4. THE NONUNIFORM TRANSMISSION LINE, **A. T. Starr.** *Ibid.*, vol. 20, June 1932, pp. 1052–63.

5. THE EXPONENTIAL TRANSMISSION LINE, **C. R. Burrows.** *Bell System Technical Journal*, New York, N. Y., vol. 17, Oct. 1938, pp. 555–73.

6. TRANSMISSION LINES WITH EXPONENTIAL TAPER, **H. A. Wheeler.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 27, Jan. 1939, pp. 65–71.

7. THE HYPERBOLIC TRANSMISSION LINE AS A MATCHING SECTION, **H. J. Scott.** *Ibid.*, vol. 41, no. 11, Nov. 1953, pp. 1654–57.

8. ORDINARY DIFFERENTIAL EQUATIONS, **E. L. Ince.** Longmans, Green and Company, London, England, 1927, pp. 23–24.

9. ORDINARY NONLINEAR DIFFERENTIAL EQUATIONS IN ENGINEERING AND PHYSICAL SCIENCE, **N. W. McLachlan.** Oxford University Press, London, England, 1950, pp. 12–13.

10. BESSEL FUNCTIONS FOR ENGINEERS. *Ibid.*, 1941.

11. APPLIED BESSEL FUNCTIONS, **F. E. Relton.** Blackie and Sons, Ltd., London, England, 1946.

12. THE REVERSION METHOD FOR SOLVING NONLINEAR DIFFERENTIAL EQUATIONS, **L. A. Pipes.** *Journal of Applied Physics*, New York, N. Y., vol. 23, no. 2, Feb. 1952, pp. 202–07.

Dynamic Behavior of a 3-Phase Magnetic Amplifier

F. J. ELLERT
ASSOCIATE MEMBER AIEE

IN MANY magnetic-amplifier applications, the transient response of the magnetic amplifier is of considerable importance. This is especially true if the magnetic amplifier is to function as an element of a feedback control system. The object of the work described in this paper was to determine experimentally the dynamic performance of a 3-phase bridge magnetic-amplifier circuit.

Polyphase magnetic-amplifier circuits, as well as single-phase circuits, exhibit both a dead-time lag (transportation lag) T_d and a time constant T_C. Measurements indicate that the polyphase magnetic amplifier also exhibits what appears to be a linear rise during the initial portion of a transient when the control circuit impedance is high. For lack of a better term, this linear rise will be called a ramp function in this paper.

The circuit diagram for the 3-phase bridge magnetic amplifier is shown in Fig. 1. The circuit consists of six identical saturable reactor units; each saturable reactor unit contains a gate winding and multiple-control and bias windings on a silicon iron core. In this investigation, cut C cores of Silectron were used; their dynamic and static hysteresis loops are shown in Fig. 2. The dynamic loop was measured under sinusoidal flux conditions at the supply frequency of 60 cycles per second.

The measured control characteristic of this circuit for the resistive load case is shown in Fig. 3 plotted in normalized form. Curve A shows the useful portion of the control characteristic and curve B (to another scale) shows the upper positive region and the entire negative region of the characteristic in greater detail. In this figure, the base for the per-unit load current i_L is the rated average load current of the circuit. The base for the per-unit magnetizing force H/H_c of the control circuit is the coercive force H_c of the dynamic hysteresis loop as given in Fig. 2.

Nomenclature

Ang = phase angle, radians
B = magnetic flux density, gausses
db = decibels
d_o = oscilloscope deflection which corresponds to the peak load current, millimeters
E_C = control voltage, volts, average
$\bar{E}_C$ = change in control voltage, volts, average
$E_i(p)$ = input voltage as a function of the Heaviside operator p
E_L = load voltage, volts, average
E_O = output voltage (I_LR_O), volts, average
$\bar{E}_O$ = change in output voltage, volts, average
$E_O(p)$ = output voltage as a function of the Heaviside operator p
e_{1-2} = voltage between line 1 and line 2, volts, instantaneous
e_{l-n} = voltage between line and the neutral, volts, instantaneous
$e_{C,A}$ = control voltage on $SR\ A$, volts, instantaneous
$e_{G,A}$ = gate voltage on $SR\ A$, volts, instantaneous
$e_i(t-T_d)$ = input voltage displaced in time by an amount T_d, volts, instantaneous
e_L = load voltage, volts, instantaneous
$e_O(t)$ = output voltage as a function of time, volts, instantaneous
$e_{R,A}$ = rectifier voltage on $REC\ A$, volts, instantaneous
f = supply frequency, cycles per second
H = component of the magnetic intensity produced by the winding, oersteds, instantaneous
$H_c = H$ for zero flux on a symmetrical dynamic hysteresis loop, oersteds, instantaneous
I_C = control current, amperes, average
I_G = gate current, amperes, average
I_L = load current, amperes, average
$\bar{I}_L$ = change in load current, amperes, average
$\bar{I}_{L,M}$ = total change in the average load current between two steady-state levels, amperes, average
i_C = control current, amperes, instantaneous
$i_{G,A}$ = gate current in $SR\ A$, amperes, instantaneous
i_L = per-unit load current
K_{AT} = ampere-turn gain
K_E = voltage gain
L_L = load inductance, henrys
L_m = log modulus (= 20 $\log_{10}$)
l_{Fe} = mean length of the magnetic circuit, centimeters
N_C = number of turns on one control winding
N_G = number of turns on one gate winding
p = Heaviside operator
$R_{B,x}$ = external resistance inserted in the bias circuit, ohms
R_C = resistance of six control windings in series, ohms
$R_{C,x}$ = external resistance inserted in the control circuit, ohms
R_G = resistance of one gate winding, ohms
$R_{G,A}$ = gate resistance of $SR\ A$, ohms

Paper 56-714, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 23, 1956; made available for printing May 3, 1956.

F. J. ELLERT is with the General Engineering Laboratory of the General Electric Company, Schenectady, N. Y.

The author wishes to express his thanks to Dr. H. F. Storm for his encouragement and assistance in the preparation of this paper.

Vol. 7
Pt. 1
1956

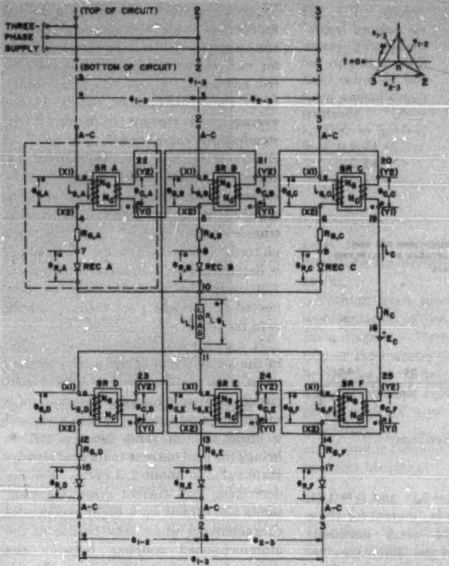


Fig. 1. Three-phase bridge magnetic amplifier circuit

R_L = load resistance, ohms
R_O = total series resistance encountered by the load current, ohms
SR = saturable reactor
T_C = control circuit time constant, seconds
T_d = dead-time lag, seconds
T_ϵ = equivalent time constant of the ramp function, seconds
T_L = load time constant (L_L/R_L), seconds
T_r = duration of the ramp function, seconds
t = time, seconds, or cycles as indicated
ω = angular frequency, radians per second
ϵ = base of natural logarithms (= 2.718 approximately)

Dead-Time Lag

Experimental Procedure

As already mentioned briefly, the response of this magnetic-amplifier circuit to a step of control voltage may consist of a dead-time lag, a ramp function, and an exponential function which is more commonly expressed in terms of a control circuit time constant. To remove the exponential portion of the transient and permit a more detailed measurement of the initial portions of the transient, a large external resistance $R_{C,x}$ was inserted in the control circuit of the magnetic amplifier. Furthermore, to reduce the effects of coupling between the control and bias circuits, a large external resistance $R_{B,x}$ was also inserted into the bias circuit. The arrangement of the control and bias windings is shown in Fig. 4. As indicated in Fig. 4(B), the bias windings are all connected in series. On the other hand, a series-parallel arrangement of windings is employed for the control circuit to permit a substantial reduction in the control circuit time constant without the need of an extremely large external resistance and a high-voltage supply. Additional measurements were also made with the control windings connected in series and without the bias windings.

Transient-response measurements were made by photographing the load current transients on an oscilloscope equipped with a triggered sweep, the triggering signal being obtained from the control circuit as switch S is closed. To ascertain the magnitude of any time constants still present in the circuit, a positive step of control voltage was applied to the control circuit with the bias circuit energized but with the magnetic amplifier disconnected from the a-c supply. Transients of the control current, bias current, and induced voltage in an additional isolated winding on one core were photographed. Both currents exhibited essentially a step function response. The induced voltage, on the other hand, exhibited exponential properties but the entire induced-voltage transient was complete in approximately 1/15 cycle of the supply frequency (60 cycles per second). This indicates a flux time constant of approximately 1/60 cycle.

If a negative step of control voltage is applied to the control circuit, the voltages induced in the gate windings are of such a polarity as to cause currents to circulate in the gate and load circuits. Thus, an effective time constant for the gate and load circuits becomes part of the transient response. This time constant does not enter the transient response to such a great degree when the control voltage step is positive because the voltages induced in the gate windings are prevented from circulating currents by the rectifiers. The effects of such coupling between the control and gate circuits will be described in more detail later.

The transient-response photographs present the instantaneous load current as a function of time. Since the load is resistive, the sixth harmonic ripple is quite predominant in the load current wave shape. In addition, ripple due to core dissymmetries is also evident. However, due to the shape of the dynamic hysteresis loop of the core material, the load current is not discontinuous at any operating point on the control characteristic.

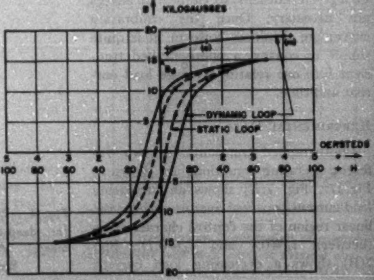


Fig. 2. Dynamic and static hysteresis loops of the magnetic cores

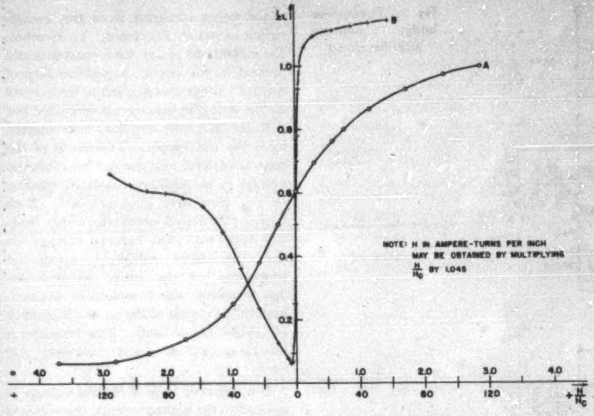


Fig. 3. Measured control characteristic plotted in normalized form

To obtain the transient response of the magnetic amplifier in terms of the average load current, it is necessary to convert the peak current values as seen on the photographed transients into average values by means of a calibration curve. Such a curve is obtained by setting the load current at a desired value and then photographing this steady-state current as displayed on the oscilloscope. When the peak load current is measured in millimeters on each photograph and related to the average load current as measured on a d-c ammeter, the calibration curve results. Two such photographs are shown in Fig. 5 and each photograph results in one point on a calibration curve such as that shown in Fig. 6, where d_o is the height in millimeters of the peak load current on the photograph. Since the peak currents for a given average load current vary from core to core, it is necessary to plot a calibration curve for each core (or each peak). Therefore, for a given transient, six calibration curves are necessary. Once the calibration curves are obtained, however, it is quite simple to convert a photographed transient into one relating average load current and time.

EXPERIMENTAL DATA

Photographed transients for two different operating conditions are shown in Fig. 7. Fig. 7(A) shows an increasing load current transient over essentially the linear region of the control characteristic (between $i_L=0.4$ and $i_L=0.7$). Fig. 7(B) shows a decreasing load current transient between $i_L=0.74$ and $i_L=0.24$. The oscilloscope gain is different for these two transients. The sixth harmonic ripple frequency and the differences between the six cores are quite evident from these photographs. In measuring these transients, the oscilloscope gain was increased to insure reasonable accuracy in the measurement of the load current changes. As a result, the zero level of the current cannot be shown on the oscilloscope screen and is actually far below the area of the screen.

An average load current transient is then obtained by measuring the peak load current at 1/6-cycle intervals on the photographed transient and relating this measurement to the average load current by means of the calibration curves. A number of such average load current transients are shown in Figs. 8 through 12 for various operating regions on the control characteristic. For these transients, I_L is the instantaneous change in the average load current from the initial steady-state value, and $I_{L,m}$ is the total change in the average load current between the initial and final steady-state values. With this notation, both increasing and decreasing load current transients take on the same form when plotted. Since the average load current is determined only at discrete intervals of 1/6 cycle, the plotted points are connected by straight lines merely to indicate the trend.

It appears from these transients that in the linear region of the control characteristic the response with a high control circuit resistance consists essentially of a maximum dead-time lag of 1/6 cycle and a linear rise (or ramp function) which brings the load current to its final steady-state value in about 2/3 cycle. For the decreasing load current cases, the transients also exhibit some slight exponential characteristics which may be due to the aforementioned coupling between the gate and control circuits.

MATHEMATICAL AND GRAPHICAL REPRESENTATION OF THE DEAD-TIME LAG

As in the case of single-phase magnetic-amplifier circuits, the dead-time lag for this 3-phase circuit arises from the sampled nature of the output. For merely the dead-time lag itself, the output $e_O(t)$ is related to the input as follows.

$$e_O(t) = e_i(t - T_d) \tag{1}$$

That is, the output is a perfect replica of

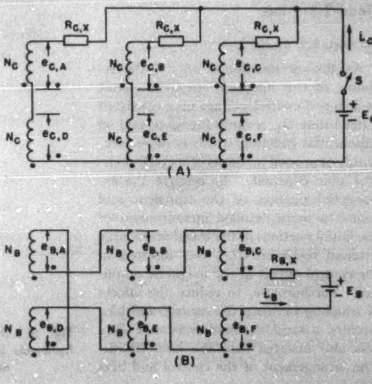


Fig. 4. Winding arrangement for dead-time lag measurements. A—Control circuit. B—Bias circuit

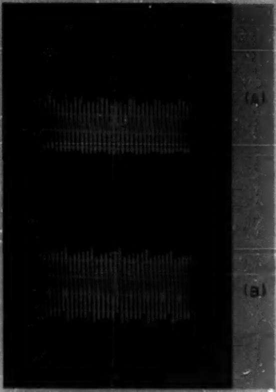


Fig. 5. Typical calibration photographs for transient response measurements. A—I_L = 2.0 amperes. B—I_L = 3.5 amperes

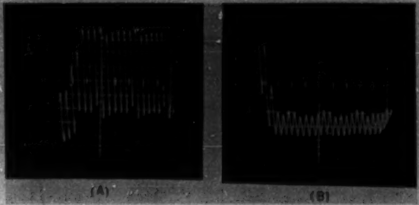


Fig. 7 (right). Typical transient response photographs for operation with high value of external control circuit resistance. A—i_L = 0.4 to i_L = 0.7. B—i_L = 0.74 to i_L = 0.24

the input but is delayed by a fixed amount T_d. The transfer function for a circuit element which produces such a delay is[1,2]

$$\frac{E_O(p)}{E_I(p)} = \epsilon^{-pT_d} = \epsilon^{-j\omega T_d} \qquad (2)$$

where p is the Heaviside operator and T_d is the dead-time lag in seconds.

In the analysis and synthesis of feedback control systems containing magnetic amplifiers, the representation of the dead-time lag in forms other than that of equation 1 becomes necessary. For example, on a Nyquist diagram (complex plane) the dead-time lag is represented by a vector of constant magnitude (unity) whose phase angle increases linearly with frequency. The expression for the phase angle is

$$\text{Ang}\,(\epsilon^{-pT_d}) = -\omega T_d \text{ (radians)} \qquad (3)$$

Therefore, on a Nyquist diagram the locus for the transfer function of the dead-time lag element is a circle of unity radius.

Perhaps even more important is the representation of the dead-time lag on a Bode diagram. Here the magnitude portion of the transfer function is described by the log modulus L_m in decibels (db) and the phase is again described by the phase angle of equation 3.[3] Since the magnitude of the transfer function is unity and is frequency invariant, the log modulus becomes

$$L_m(\epsilon^{-pT_d}) = L_m(1) = 0 \text{ db} \qquad (4)$$

Therefore, to add the transfer function for a dead time lag to the existing transfer function of a system, it is merely necessary to add the phase angle of the dead-time lag as given in equation 3 to the existing phase angle of the system. As indicated in equation 4, no change is necessary in the amplitude portion of the system transfer function. Although the phase angle as given in equation 3 is a linear function of the angular frequency, it will not appear as a straight line on a Bode diagram because of the logarithmic scale which is commonly employed for the angular frequency ω.

Whenever an analogue computer study is made of a system containing a dead-time lag element, it is desirable to represent the transfer function of the dead-time lag element by means of physically realiz-

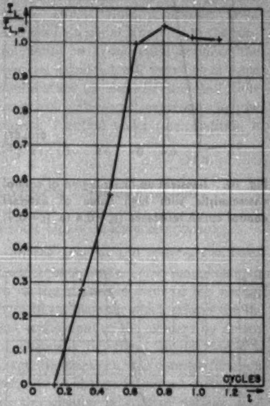


Fig. 8. Small transient up linear region of control characteristic with high value of external control circuit resistance (i_L = 0.5 to i_L = 0.6)

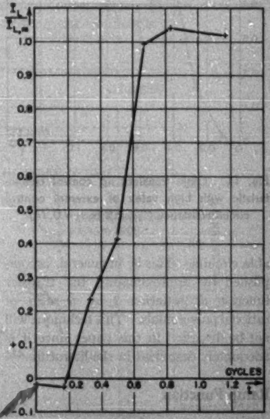


Fig. 9. Transient up lower region of control characteristic with high value of external control circuit resistance (i_L = 0.2 to i_L = 0.4)

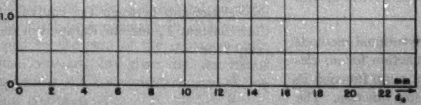


Fig. 6. Calibration curve relating the average load current and oscilloscope deflection corresponding to peak load current

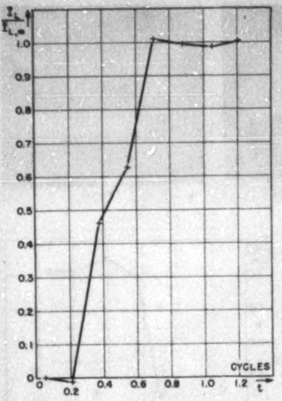


Fig. 10. Transient up linear region of control characteristic with high value of external control circuit resistance ($i_L=0.4$ to $i_L=0.7$)

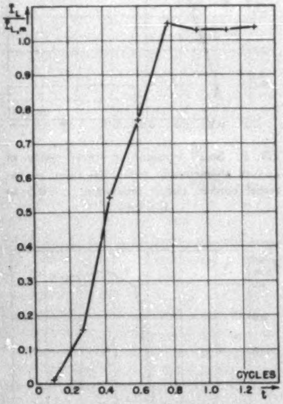


Fig. 11. Large transient up control characteristic with high value of external control circuit resistance ($i_L=0.2$ to $i_L=0.74$)

able circuits. This is, in general, accomplished by approximating the transfer function of equation 2 by a ratio of rational polynomials. This technique will not be discussed in this paper since it is adequately described in the literature.[4,5,6]

Ramp Function

Experimental Data

The transients shown in Figs. 8 through 12 clearly indicate the presence of a somewhat linear rise in the response when a step of control current is applied to the control windings. The few transients shown here are representative of the many transients which were actually measured. It is interesting to note that for the increasing load current transients, the duration of the linear rise is approximately 1/2 cycle of the carrier frequency. Experimentation has shown that the duration of the ramp is not a constant for all operating conditions. If the a-c supply voltage is decreased, resulting in underexcitation of the magnetic amplifier, the duration of the ramp increases. For a-c supply voltages from slightly below nominal voltage to voltages above the nominal value, the duration of the ramp remains essentially constant at the value shown in Figs. 8 through 12. The ramp appears to be caused by an accommodation process in the cores whereby the cyclical core flux adjusts itself from operation on one minor hysteresis loop to another, as dictated by the magnitude of the control current step. The exact nature of the flux adjustment in each core is a function of its operating point on the hysteresis loop at the instant the step is applied. For cores which are saturated at that time no flux adjustment can occur.

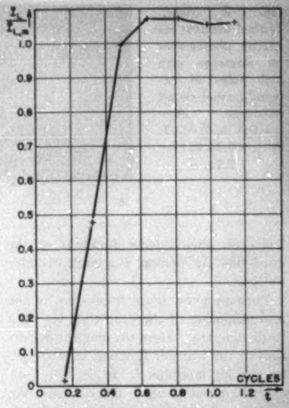


Fig. 12. Large transient down control characteristic with high value of external control circuit resistance ($I_L=0.74$ to $i_L=0.24$)

Fig. 13. Transient up linear region of control characteristic with intermediate value of external control circuit resistance ($i_L=0.4$ to $I_L=0.6$) $T_C=0.0325$ second

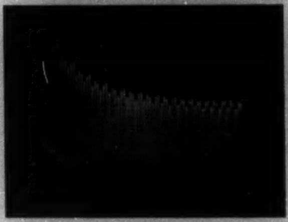


Fig. 14. Transient down linear region of control characteristic with intermediate value of external control circuit resistance ($i_L=0.6$ to $i_L=0.4$) $T=0.05$ second

Mathematical Representation of the Ramp Function

It is possible by conventional methods to obtain a transfer function for an element which produces a ramp function in response to an applied step function. The expression for the transfer function of such an element is

$$\frac{E_O(p)}{E_i(p)}=\frac{1}{pT_r}(1-\epsilon^{-pT_r}) \qquad (5)$$

where T_r is the duration of the ramp function in seconds. For this 3-phase circuit, the duration of the ramp function is approximately 1/2 cycle of the supply frequency f for increasing load current transients. Therefore, for this case, equation 5 becomes

$$\frac{E_O(p)}{E_i(p)}=\frac{2f}{p}\left(1-\epsilon^{-\frac{p}{2f}}\right) \qquad (6)$$

In some cases, it may be desirable to represent the ramp function by a transfer function which is of necessity simpler and, as a result, less accurate. One possibility is to represent the ramp function by an equivalent single time constant. If the time integral of the instantaneous error between the values of the ramp and the equivalent exponential is to be zero, the relationship between the equivalent time constant T_e and the duration of the ramp T_r is

$$T_e=\frac{T_r}{2} \qquad (7)$$

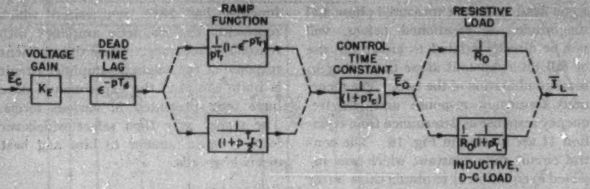


Fig. 15. Block diagram of 3-phase bridge magnetic-amplifier circuit for increasing load current transient

The transfer function for the element producing the ramp then becomes

$$\frac{E_O(p)}{E_i(p)}=\frac{1}{(1+pT_e)}=\frac{1}{\left(1+p\frac{T_r}{2}\right)}. \qquad (8)$$

It should be clear that equation 8 describes only approximately the transient behavior of the ramp.

Time Constant

EXPERIMENTAL PROCEDURE

For low or intermediate values of control circuit resistance, the exponential function becomes the predominant part of the transient response. The evaluation of the control circuit time constant from the measured transients is then quite simple.

For these measurements, the six control windings were connected in series as shown in Fig. 1; the bias windings were not used. No time constant measurements were made with the series-parallel arrangement of control windings shown in Fig. 4(A). The calibration and measurement techniques described in conjunction with the dead-time lag were also followed here. However, in this case, the total control circuit resistance was reasonably small.

EXPERIMENTAL DATA

Two photographs of load current transients over the essentially linear region of the control characteristic are shown in Figs. 13 and 14. Fig. 13 shows an increasing load current and Fig. 14, a decreasing load current. The values of the measured time constants are given below the respective figures. As indicated, the time constant of the decreasing or falling load current is greater due to the coupling between the gate and control circuits. The total time constant in this case is equal to the sum of the control circuit time constant (as given by the rising transient) and an equivalent gate circuit time constant.

MATHEMATICAL REPRESENTATION OF THE TIME CONSTANT

The control circuit time constant T_C, for the case of the series-connected control windings, may be calculated with reasonable accuracy from the following equation

$$T_C=\frac{1}{2f}K_{AT}\frac{R_O}{R_C}\left(\frac{N_C}{N_G}\right)^2 \text{ (seconds)} \qquad (9)$$

where K_{AT} is the ampere-turn gain of the magnetic-amplifier circuit, R_O is the total resistance encountered by the load current, and R_C is the total control circuit resistance. Equation 7 is identical to the expression for the control circuit time constant of a single-phase magnetic amplifier.[7] The measured time constants for increasing load currents were within a few per cent of the values calculated from equation 9 when R_O was corrected to include the effects of the aftersaturation inductance of the gate windings. This is certainly within the required engineering accuracy.

The transfer function for a circuit element which produces an exponential output in response to a step function input is[8]

$$\frac{E_O(p)}{E_i(p)}=\frac{1}{(1+pT)} \qquad (10)$$

where T is the time constant of the exponential expressed in seconds. For an increasing load current transient, T becomes the control circuit time constant T_C; for a decreasing load current transient, T becomes the sum of the gate and control circuit time constants. If a sinusoidal a-c signal is applied to the control circuit of the magnetic amplifier, the response exhibited will result from some combination of the rise and fall time constants. The exact expression for the transfer function under this condition is not known.

Complete Transfer Function

BLOCK DIAGRAM REPRESENTATION

In the analysis and synthesis of feedback control systems, it is quite common to represent each element of the system by a block diagram which relates the input and output of the element by means of a transfer function. In the case of a magnetic amplifier, this transfer function relates the change in output voltage $\overline{E}_O$ to the change in control voltage $\overline{E}_C$.[9] For the resistive load case, the change in output voltage $\overline{E}_O$ is very simply related to the change in load current $\overline{I}_L$ by means of the output resistance R_O. The transfer function of the 3-phase bridge magnetic-amplifier circuit for an increasing load current transient is

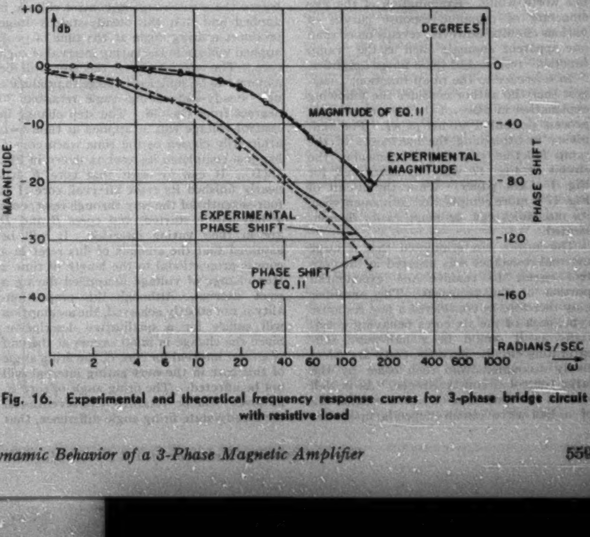


Fig. 16. Experimental and theoretical frequency response curves for 3-phase bridge circuit with resistive load

$$\frac{E_O}{E_C}=K_E\left(\epsilon^{-\frac{p}{6f}}\right)\frac{2f}{p}\left(1-\epsilon^{-\frac{p}{2f}}\right)\frac{1}{(1+pT_C)} \tag{11}$$

where K_E is the voltage gain of the magnetic-amplifier circuit, the assumed dead-time lag T_d is $1/6f$ seconds (1/6 cycle), and the duration of the ramp function is 1/2 cycle. The general block diagram is shown in Fig. 15. The two possible representations of the ramp function are also shown. Depending on the nature of the load, one or the other of the blocks shown at the output must be used to obtain the change in load current. For the inductive, d-c load case, T_L is the load time constant L_L/R_L in seconds.

FREQUENCY RESPONSE MEASUREMENTS

To verify experimentally the transfer function equation 11, the frequency response of the magnetic amplifier was measured by applying a sinusoidal voltage of adjustable frequency to the control circuit and measuring the magnitude and phase of the output voltage. Since the control current under these conditions is sinusoidal and causes positive and negative excursions of the load current about some fixed value, the measured phase and magnitude, as mentioned before, will not be directly related to either the rise or fall time constant alone but rather to some combination of the two. The measured frequency response and the frequency response as determined from equation 11 are plotted in Fig. 16. The control circuit time constant which was inserted in equation 11 to obtain these latter frequency response data is equal to the average of the rise and fall time constants as measured in the step function tests. For the curves of Fig. 16, the voltage gain was assumed to be unity; for this reason the magnitude approaches 0 db at low frequencies. The correlation between the theoretical and the experimental data is quite good and justifies the use of equation 11 to represent the 3-phase bridge circuit.

Conclusions

The emphasis in this paper has been on the presentation of experimental data which describes the transient response of the 3-phase bridge magnetic-amplifier circuit. From these data, a transfer function has been synthesized which represents the magnetic amplifier with reasonable accuracy. Since the transient response of the magnetic-amplifier circuit is quite complex, the experimental data have been presented in various forms. The reader may then select whichever form is most familiar to him and best meets his needs.

References

1. Transients in Linear Systems, **M. F. Gardner, J. L. Barnes.** John Wiley & Sons, Inc., New York, N. Y., 1942, p. 236.

2. Magnetic Amplifiers, **H. F. Storm,** *et al. Ibid.*, 1955, p. 229.

3. Principles of Servomechanisms, **G. S. Brown, D. P. Campbell.** *Ibid.*, 1948, p. 237.

4. Sur la Réprésentation approchée d'une function par des fractions rationnelles. *Annales*, École Normale, Paris, France, vol. 9, no. 3, 1892, pp. 1–93.

5. Time Domain Approximation by Use of Pade Approximants, **R. D. Teasdale.** *Convention Record*, Institute of Radio Engineers, New York, N. Y., 1953, pt. V, pp. 89–94.

6. Transport Delay Simulation Circuits, **C. H. Single, G. S. Stubbs.** *WAPD-T-38*, Westinghouse Electric Corporation, Pittsburgh, Pa.

7. **Storm,** *op. cit.*, p.269.

8. **Storm,** *op. cit.*, p. 146.

9. **Storm,** *op. cit.*, p. 228.

◆

Discussion

G. E. Lynn (Westinghouse Electric Corporation, Baltimore, Md.): This paper is a welcome addition to the literature on a somewhat neglected portion of the magnetic-amplifier field. The approximate solutions to the dynamic response have been so widely accepted that most investigators have apparently assumed that further study is not worth while. Examination of the fine structure of dynamic-response curves of various circuits, however, reveals more than one apparent anomaly such as the "ramp function" response of the 3-phase amplifier.

In reference to the ramp function, I suggest that the author consider the following explanation in place of the accommodation process presented in his paper. For simplicity in explaining the mechanics of the ramp function response phenomenon, the circuit of Fig. 17 will be substituted for Fig. 1 of the paper. While the circuit of Fig. 17 is more complicated componentwise, its mechanics can be more easily demonstrated.

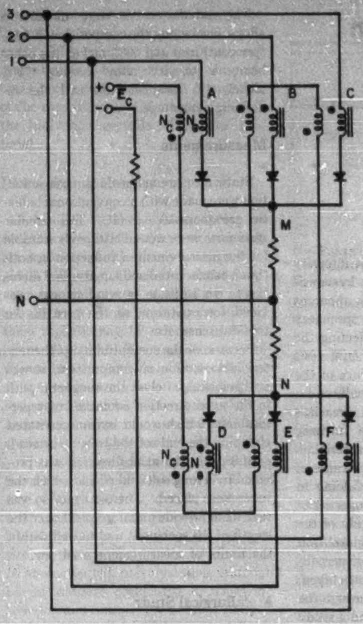
The author has stipulated that a large external resistance was inserted in the control circuit to remove the exponential portion of the transient. The amplifier may therefore be considered a fast-response type, each of the six cores behaving essentially as it would in a half-wave self-saturating circuit, since the cores are effectively decoupled from each other by the large control circuit resistance. As is well known, the dead-time or transportation lag of a half-wave circuit depends upon the instant of time in the cycle of supply frequency at which the step change of input voltage is applied.

The circuit under discussion consists of six cores and the phase relations are such that a reset period of one core begins every 60 electrical degrees, as shown in Fig. 18(B). The rectified phase to neutral voltages are shown for reference in Fig. 18(A). For the purpose of explanation, it will be assumed that a steady state control voltage exists before the step change in control voltage is applied and that this steady-state voltage produces a firing angle at the time of peak applied voltage in the gating interval of each core. The magnitude of the step will be assumed to be sufficiently large to produce a new steady-state firing angle retarded 30 degrees from the old. The step change in control voltage will be applied at time $t=t_0$ arbitrarily chosen as the time when core C has just completed its reset as shown in Fig. 18(B). It can be seen that core E has nearly finished its reset interval, core A is four-sevenths of the way through reset, core F has barely started, and cores B and D are in their gating intervals. It will be assumed that the amount of flux reset in a core is proportional to the length of time a step change of voltage is applied during a reset interval. Although this proportionality is not strictly achieved, the assumption will suffice for a qualitative description. Since the change in input occurs at the end of a reset interval for core C, the firing angle of this core in the next gating interval will not be affected. The firing angle of core E, however, will be retarded two-sevenths of the steady-state firing angle difference, that of core A four-sevenths, etc. The output voltage as viewed across points M and N of Fig. 17 will be as shown in Fig. 18(C). Steady-state is seen to be achieved at time $t=t_1$, 11/12 cycle after the application of the step change in input voltage. Dead time in this case is a little over 1/3 cycle. Had the step input been applied 1/12 cycle earlier or later, the dead time would have been 1/6 cycle and the transient would have been completed in 5/6 cycle. The latter figures are seen to be in good agreement with the author's measured values of 0.17 cycle dead time and 0.7 to 0.8 cycle for full response.

The existence of the ramp function response would thus be explained as being caused not by "an accommodation process in the cores whereby the cyclical core flux adjusts itself from operation on one minor hysteresis loop to another," but as existing simply because six cores must change their firing angles in the length of time reserved for two cores in a single-phase full-wave circuit or for one core in the half-wave circuit.

While it is true that a path for possible neutral currents exists in the circuit of Fig. 17, it would appear from the 3-phase symmetry of the circuit of Fig. 1 that the transient behavior of the two circuits would be essentially similar. For this reason it is felt that the foregoing explanation of the transient response is valid for both.

F. J. Ellert: The author appreciates Mr. Lynn's contribution in discussing in detail



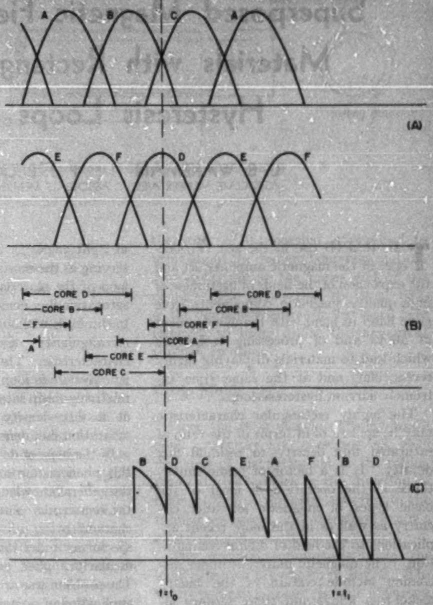
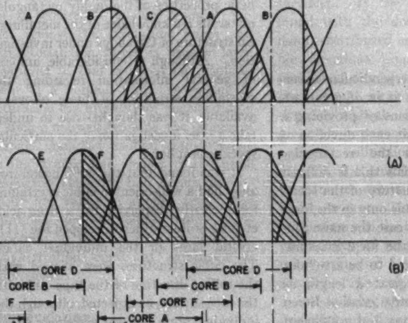


Fig. 17 (above left). Revised circuit for discussion

Figs. 18 (above right) and 19 (below left)

A—Reference voltage wave shapes
B—Resetting intervals
C—Output voltage transient



the cause of the ramp function. The details were also brought out during the author's oral presentation of his paper and it is gratifying to record agreement on this part of the discussion.

It will be seen from Fig. 18(B) that three cores (E, A, and F) will traverse minor hysteresis loops which are intermediate, namely between the steady-state loops before and after the transient. The term accommodation is used to indicate the existence of intermediate hysteresis loops.

The output voltage as viewed across points M and N, and shown in Fig. 18(C) is in error. (There are two different points in Fig. 18(C), both being designated by N. One N designates the transformer neutral, and the other designates the star point of three rectifiers. To clarify the following, the transformer neutral will be designated by N'.) The output voltage of the circuit shown in Fig. 17 is the sum of the voltages between MN' and $N'N$. Using exactly the same premises as Mr. Lynn, the correct output voltage (between M and N) is shown in heavy lines in Fig. 19(C). It will be readily recognized that Mr. Lynn's output voltage of Fig. 18(C) is deficient by the amounts shown in shading in Fig. 19(C).

Superposed Magnetic Fields in Materials with Rectangular Hysteresis Loops

C. B. WAKEMAN
ASSOCIATE MEMBER AIEE

F. J. BECK
ASSOCIATE MEMBER AIEE

THE RAPID rise in the last two decades of the magnetic amplifier art and the expansion of the fields of usefulness of high-quality magnetic components have gone hand in hand with the development of alloys and of processing techniques which lead to materials displaying highly rectangular, and at the same time extremely narrow, hysteresis loops.

The highly rectangular characteristic may be spoken of in terms of the ratio of saturated flux density to residual flux density. It is a factor of prime importance in the operation of most of the basic types of magnetic amplifier circuitry as well as in the more recent applications to the field of digital computation. The magnetic materials under discussion include certain of the binary nickel-iron alloys and those ternary alloys which include small percentages of molybdenum. These alloys are of interest not only because of their utility as practical circuit elements. Their peculiar crystalline properties provide the means of verifying and extending many of the theories of magnetic action which have been developed on the basis of experiments with single crystals.

In addition to their remarkable hysteresis properties, these materials exhibit interesting characteristics when subjected to simultaneously applied perpendicular fields. The relative simplicity of fabrication of the polycrystalline and grain-oriented specimens has permitted an experimental investigation of these characteristics to be carried out in conjunction with an analytical study.

The application of a uniform fixed field at right angles to the magnetic intensity serving as the excitation for the hysteresis loop of the material causes the apparent saturation flux density of these specimens to diminish without greatly affecting the rectangularity of the new hysteresis characteristic. The coercive force of the new hysteresis loop remains small, and a relatively sharp saturation level is reached at a flux density below the intrinsic saturation flux density of the material.

In the hope of determining the cause of this phenomenon and thereby adding to the general knowledge of the properties of the rectangular materials, a study of the characteristics of selected nickel-iron specimens under the influence of perpendicularly applied fields was undertaken. The problem was approached through the application of domain theory and a study of the crystalline properties of the magnetic materials.

Specimens

In any study of perpendicular superposed magnetic fields it is of great advantage to find a means of providing a closed magnetic path in each direction of measurement. With the rectangular hysteresis loop materials this is difficult because the physical nature of the specimens which are available only in the form of thin tapes. In this case the most convenient approach to the ideal measurement geometry appeared to be a tubular specimen having as great a length as possible, a small, effective cross-sectional area in the axial direction, but a sufficient number of circumferential tape wraps to provide good magnetic closure in the toroidal sense. Samples were obtained having the geometry outlined, with dimensions as shown in Fig. 1. In each specimen, ten wraps of tape having a thickness of approximately 51 microns were wrapped on a ceramic tube. The rolling direction corresponds to the circumferential direction around the tube. Each of the cores was placed on the tube prior to the final heat treatment appropriate to the material.

The tubular cores were nickel-iron alloys, one having the composition of 50% (per cent) iron and 50% nickel, the other being a ternary alloy having 79% nickel, 4% molybdenum, and the remaining percentage iron.

Measurements

Static measurements of magnetic induction were made with a conventional ballistic galvanometer circuit. The tubular specimens were wound with coils suitable for the measurement of induction in both the circumferential and longitudinal directions. In addition, a winding was provided for excitation of the core in the toroidal sense.

Because of the susceptibility of the materials to torsion or compression, it was not practical to close the magnetic path in the axial direction by means of a permeameter which would have necessitated clamping the ends of the tubes. Instead, excitation in the axial direction was provided by a long solenoid within which the tubes were placed. The construction was such as to provide a uniform field over the length of the specimen, and to well within the limits of measurement accuracy.

Metallurgical Study

Fundamental to a study of the magnetic properties of the highly rectangular materials is a knowledge of the metallurgical structure of the alloys under investigation. Although a considerable amount of general information regarding the crystalline structure of these alloys is available, it was thought wise to undertake a brief investigation of the particular specimens at hand.

It was found that the 50% nickel-iron alloy had a well-oriented cubic crystalline form clearly displayed by a deep-acid-etching process which developed the (111) crystal faces of the individual cubes. The cubic structures were aligned with the rolling direction of the tape and with the tape surface, as expected, although the individual grain size was extremely small. X-ray transmission patterns were also taken. These confirmed the findings that the structure consists of extremely small grains, each of which is well oriented to the co-ordinates of the tape. The easy directions of magnetization for an alloy of this composition are along the cube edges, and thus the easy directions of magnetization for the tape as a whole are parallel to the rolling direction, and across the rolling direction.

X-ray transmission patterns of the crystal structure of the *4-79* molybdenum

---

Paper 56-717, recommended by the AIEE Basic Sciences Committee and Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 27, 1956; made available for printing May 22, 1956.

C. B. Wakeman is with Magnetics, Inc., Butler, Pa., and F. J. Beck is with Yale University, New Haven, Conn.

This paper is part of a dissertation presented by C. B. Wakeman for the degree of Doctor of Philosophy at Yale University.

permalloy showed the Debye rings to be typical of a completely unoriented specimen. This predicts a nondirectional magnetization characteristic for this material because of the random orientation of the easy directions of magnetization of the individual crystals within the material.

Discussion of Results

According to the domain theory of magnetization, all ferromagnetic materials are made up of completely saturated regions. Within a given region the magnetic moment vectors of all of the atoms are aligned with one another; yet each of these regions may be oriented in a different direction. The magnetization of the material as a whole is the vector sum of the magnetizations of the individual regions or domains. The average magnetization, when no external field is applied, may well be zero.

The magnetization process in the material may be one of rotation of the intensity of magnetization vectors of the domains, or it may be a process of domain boundary motion changing the relative size of the variously oriented domains. In an unoriented, polycrystalline material having nonzero coefficients of magnetostriction and anisotropy, a major portion of the energy required to saturate the material goes into domain rotation. The energy required to produce domain rotation against the crystalline forces and magnetostriction forces is usually large compared to that required for displacement of a domain wall. In raising to saturation the intensity of magnetization of such a magnetic sample, the process of domain boundary motion corresponding to the straight-line portion of the usual magnetization curve first occurs at relatively low values of applied field. The rotational process follows later, corresponding to the curved portion of the magnetization curve as saturation is approached.

Nomenclature

H = total magnetizing force
H_T = total effective transverse magnetizing force (in the rolling plane, normal to the rolling direction)
H_{Ta} = applied transverse magnetizing force
H_{Ro} = magnetizing force parallel to the rolling direction
B_T = flux density in the rolling plane, normal to the rolling direction
I_{Ro} = intensity of magnetization, parallel to the rolling direction
I_T = transverse intensity of magnetization, in the rolling plane, normal to the rolling direction
I = intrinsic intensity of magnetization
N = demagnetizing factor
E = total crystal energy
E_A = energy of crystal anisotropy

Electromagnetic units are used throughout the paper.

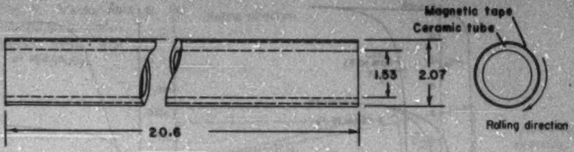


Fig. 1. Dimensions of tubular cores (centimeters)

50% Nickel-Iron Alloy

Consider now a magnetic material from which the randomness of the crystalline structure has been removed, and each of the cubic elements of the lattice is fairly closely aligned with its neighbor. This situation has been shown to exist in the 50% nickel-iron specimen under investigation. If the cubic lattice is arranged, as has been indicated, with the cube edge lying parallel to the rolling direction, it should be possible to shift the entire saturation flux solely through the process of domain boundary motion, as in the case of a properly oriented, single crystal specimen. If the material is initially found with the domain directions randomly distributed in each of the six possible directions, applying a small field parallel to any of the crystal-axis directions should serve to align each of the domains with the field. Since the field direction corresponds to an easy magnetization direction for each of the crystals, the anisotropy energy required to obtain complete field alignment would ideally be zero. The magnetization curve in this case will rise to the saturation value of intensity of magnetization with very small values of fields, and will saturate sharply. The value of applied field required to bring about the domain boundary motion is dependent on a number of factors, which include impurities and local strain effects.

In the 50% nickel-iron tapes the crystal orientation is not perfect, and strain removal has not been complete. The major portion of the magnetization takes place, however, with very small fields applied, and the saturation is quite abrupt. These features are illustrated by the static hysteresis loop for the tubular specimen of this material, as shown in Fig. 2. The straight sides of the loop show that a uniform magnetizing force of the order of 0.1 oersted is required to bring about the domain boundary motion. The curvature at the upper end of the characteristic shows that the final 10% of the domains requires an increased field strength to bring about the alignment through rotation.

Cross Fields

Since the 50% nickel-iron alloy has a cubic structure, it possesses an easy direction of magnetization in the rolling plane at right angles to the direction of rolling. This is the axial direction in the tubular specimens and, therefore, is not a closed magnetic path.

The magnetic characteristics in the axial direction are shown in Fig. 3. The sharp saturation point is still apparent, but, in order to obtain from these data the true hysteresis loop of the material in the axial direction, it would be necessary to know the demagnetizing factor for the particular tube geometry. Furthermore, if good results are to be obtained, the demagnetizing factor must be known fairly accurately. The effective field strength acting axially on the tube is the difference between the external applied field strength of the solenoid and the field due to the surface density of magnetic charge induced at the ends of the specimen. Expressed quantitatively

$$H_T = H_{Ta} - NI_T \tag{1}$$

The transverse intensity of magnetization, however, is a function of $\mathbf{H}_T$ and $\mathbf{B}_T$.

$$B_T = H_T + 4\pi I_T \tag{2}$$

So

$$H_T = H_{Ta} - \frac{N}{4\pi}(B_T - H_T) \tag{3}$$

A calculation of demagnetizing factor may be made by using the curves of Fig. 3. If the magnetizing force required for small values of transverse induction is considered negligible, it is possible to find the factor which will cause the magnetization curve for the 50% nickel-iron tube to lie vertically along the B_T axis, for value of

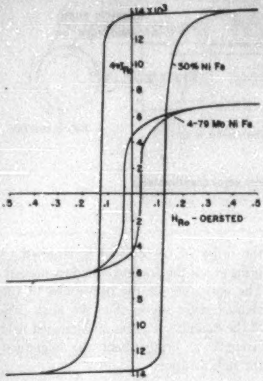


Fig. 2. Static characteristics of tubular cores in the circumferential direction

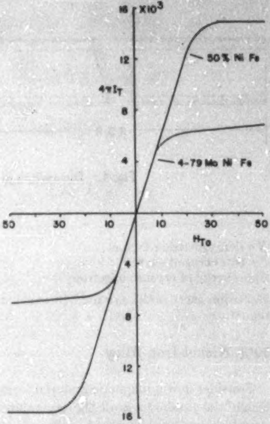


Fig. 3. Axial magnetization characteristics of the tubular cores

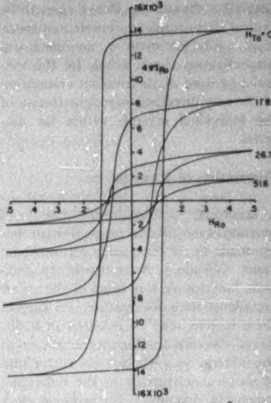


Fig. 4. Hysteresis loops of 50% nickel-iron alloy tube with various applied transverse fields

flux density less than about 8.0 kilogauss. Making use of equation 1, N is found by trial to be

$$N = 0.02 \tag{4}$$

The demagnetizing field is caused by the surface distribution of magnetic charge induced at the ends of the specimen, and the demagnetizing factor is properly so-called only when the induced charge arises from a uniform intensity of magnetization within the sample. This happens with a spheroidal configuration, since the magnetization is then uniform throughout the material, but not in a rod or a tubular sample. It should be recognized that any discussion of a demagnetizing factor as a constant is not accurate for a rod on tubular specimen. Examining the properties of the material under the influence of superposed fields, however, should provide further information on this point, as well as on the manner in which the magnetic changes take place.

Superposed Fields

Consider that situation which exists when two fields are superposed at right angles in the 50% nickel-iron tubular specimen. As has been shown, each field is oriented to lie along a direction of easy magnetization. By first applying a fixed field, H_{Ta}, along the axis of the tubular specimen, a sufficient number of domains will be aligned with the field in the axial direction to cause a demagnetizing field which is very nearly equal, and opposite, to H_{Ta}. Since the hysteresis loop width for all flux levels below saturation is a small and essentially constant quantity, and since the demagnetizing factor is relatively large, the number of domains aligned in the axial direction will be very nearly proportional to the applied field. We can write from equation 1 when $H_T \ll H_{Ta}$

$$H_{Ta} = NI_T \tag{1A}$$

If the transverse induction were slightly reduced by the action of the superposed field H_{Ro}, the effective transverse field would rise sharply, tending to return these domains to the transverse direction. The remaining domains, however, would be free to be acted upon by the field in the rolling direction, and for small values of H_{Ro} would be essentially independent of H_T. The effect is much as though a part of the magnetic material had been physically removed, with the result that the total saturation flux is lowered, but the requirements of the magnetizing field in the rolling direction are left unchanged.

The characteristics of the tubular sample illustrate this action. In Fig. 4 hysteresis loops of B_R versus H_{Ro} are plotted for selected values of applied transverse field. As the transverse field is increased, the apparent saturation flux density in the rolling direction decreases. The rectangularity of the loop is fairly well maintained, however, until the transverse field becomes so large that saturation is reached, and $I_T \cong I$. Beyond this point, there should be no free domains which are not under the direct influence of the transverse field.

Case I. $H_{Ta} > NI$

In this case a large transverse field is applied along the tube axis, so that all of the domains are caused to align with the cross-rolling direction of the tape. In order for this to be true, according to equation 1, H_{Ta} must be larger than NI, and the effective transverse field is given by

$$H_T = H_{Ta} - NI \tag{1B}$$

If a field H_{Ro} is then applied in the rolling direction there will be a net field acting on each of the crystals which is not parallel to one of the crystal axes. Referring to Fig. 5, the total magnetizing field H is the vector sum of H_{Ro} and H_T, and lies at an angle θ_3 which depends upon the relative magnitude of these components. It should be possible to calculate, using energy relations, the angle of domain rotation θ_2 resulting from the application of these fields. The rotation angle must represent a condition of minimum crystal energy, which, if the energy of magnetostriction may be considered negligible, is given by the sum of the field energy and the anisotropy energy. The anisotropy energy may be expressed in terms of the cosines of the angles between the field and the cubic crystalline directions

$$E_A = K_0 + K_1(\alpha_x^2\alpha_y^2 + \alpha_y^2\alpha_z^2 + \alpha_z^2\alpha_x^2) + K_2\alpha_x^2\alpha_y^2\alpha_z^2 + \ldots \tag{5}$$

When this result is simplified for the particular geometry, and the energy of the field is added, the expression for the total energy becomes

$$E = K_0 + K_1 \sin^2\theta_2 \cos^2\theta_2 - HI\cos(\theta_2 - \theta_3) \quad (6)$$

If the angles are related by the use of equation 1 and Fig. 5

$$H_{Ro}\cot\theta_3 = H_{To} - NI\cos\theta_2$$

and

$$H = \frac{H_{Ro}}{\sin\theta_3}$$

With these substitutions equation 6 reduces to

$$E = K_0 + K_1 \sin^2\theta_2 \cos^2\theta_2 - H_{Ro}I \times \left[\cos\theta_2\left(\frac{H_{To} - NI\cos\theta_2}{H_{Ro}}\right) + \sin\theta_2\right] \quad (6A)$$

If this energy expression is differentiated with respect to θ_2 and the resulting expression is set equal to zero

$$\frac{K_1}{2}\sin 4\theta_2 + H_{To}I\sin\theta_2 - H_{Ro}I\cos\theta_2 - NI^2\sin 2\theta_2 = 0 \quad (6B)$$

Solving this equation will show the presence of a single maximum and minimum energy condition for the intensity of magnetization vector in the range $0<\theta_2<\pi/2$. The solution representing the minimum energy point can be chosen by inspection of the vector diagram of Fig. 5. For the situation in which $H_{Ro}<H_T$, the value of θ_2 representing the minimum of crystal energy must lie in the range

$$0<\theta_2<\tan^{-1}\frac{H_{Ro}}{H_T} \quad (7)$$

In order to make use of equation 6(B) for numerical computation, it is necessary to assign values to the constants N, I, and K. The demagnetizing factor N will be assumed to be equal to the measured value of Equation 4 as a first approximation.

K_1 is the first anisotropy constant and is a positive number for nickel-iron alloys of less than about 73% nickel content. Bozorth and Walker[1] have shown values ranging in the neighborhood of 7×10^3 ergs/cm^3 (per cubic centimeter) for the 50% alloy, while an earlier summary by McKeehan[2] gives the value as 3.3×10^4. This wide range is indicative of the variations to be expected with different methods of handling. The later results of Bozorth and Walker are probably more nearly representative of the tape samples, and the value $K_1=10^4$ ergs/cm^3 will be used for computation.

It is possible with these values to calculate the expected rotation angles for any given applied longitudinal and circumferential fields. The components of the intensity of magnetization in the two directions may then be obtained and compared with the values measured with the tubular specimens. This comparison was made, and it was found that, while the slopes of the calculated curves agreed quite well, the calculated value always fell below the measured points. For each of the measured curves, it appears that the material is not entirely saturated in the transverse direction in spite of the large value of applied field. A small group of domains appears to remain independent of the transverse field and is easily influenced by circumferential fields of the order of the coercive force of the material.

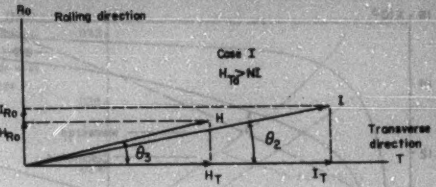


Fig. 5. Vector relationships for superposed fields with H_{To} large

The reason for the independent action of this small group is not clear. It appears that in this group there are additional factors in the energy relations which have not been accounted for. One factor might be a magnetostrictive effect caused by local stresses within the material; but, as will be shown, this explanation is not adequate for all the materials under investigation since they have widely different magnetostrictive constant although the measured effect is much the same in all cases.

If this deviation from the calculated values of I_{Ro} is accounted for by assuming that such a group of independent domains exists, the remainder of the magnetization process appears to take place, much as has been proposed, by a process first of alignment of the major proportion of the domains with the transverse axis and then a rotational process out of the transverse axis as H_{Ro} is increased from zero. The magnitude of the independent domain group may be estimated by extrapolating the curves for the various values of H_{To} back to the axis of ordinates.

If the values of I_{Ro} are recomputed by removing the group of independent domains from the calculations and adding it arithmetically to the value obtained from equation 6(A), the curves fall as shown in the lower portions ($26.7<H_{To}\leqslant101$) of Fig. 6.

As the value of the transverse field is decreased, a point will be reached at which the effective transverse field and the circumferential field will be equal. At this point $\theta_2=\pi/4$. In terms of the properties of a single cubic crystal, any further decrease in the axial field should lead to a minimum energy condition with all of the domains oriented more closely with the circumferential direction than with the axial direction.

If the domains attempt to shift to this lower energy condition, however, the demagnetizing field in the transverse direction will decrease, causing a corresponding increase in the effective transverse magnetizing force. The increased H_T will in turn reduce the angle made by the total magnetizing vector with the T axis. This reduction in θ_2 must reach an equilibrium point at $\theta_2=\pi/4$. Below this value, H_T is larger than is consistent with the value of demagnetizing field. For θ_2 larger than this value, all domains must lie near the rolling direction axis, making H_T too large. If $\theta_2=\pi/4$, however, the transverse direction and the rolling direction represent conditions of equal crystal energy. Thus the domains can divide, with a part of the total lying nearly parallel to each of the axes.

Under these conditions, the magnetization changes taking place cannot be divided sequentially, with the domain boundary motion occurring first, and the rotation taking place later as the applied field is increased. Here the domains will shift alignment between the rolling direction and the transverse direction by a process of 90-degree boundary wall displacement, and, at the same time, both groups of domains will be rotated out of exact alignment with the respective axes by the action of the total effective field.

To calculate the conditions of minimum energy in the crystal, it is necessary, therefore, to account for the field energy and the anisotropy energy of each of these groups of domains.

Case II. $H_T=H_{Ro}$

If H_{Ro} is increased beyond the value corresponding to the coercive force of the material, domains will be removed from the transverse axis to align with H_{Ro}. However, the demagnetizing field will grow until an equilibrium is reached at $H_T=H_{Ro}$. The conditions will then be represented by the vector diagram of Fig.

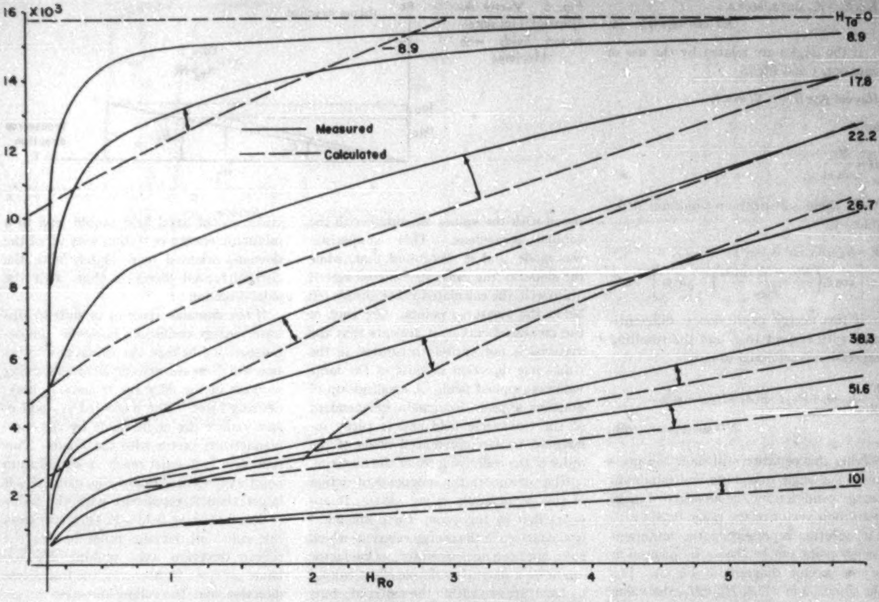


Fig. 6. Comparison of calculated and computed magnetization curves for the 50% nickel-iron alloy tube

7. The two groups of domains are represented by I_1 and I_2 respectively, and make angles of θ_1 and θ_2 with the rolling and transverse directions.

The energy expressions may be written for each of these groups of domains if an assumption is made regarding the value of the anisotropy constant for a group of domains which represent less than the saturated value of magnetization. For lack of better information, it has been assumed that the anisotropy constant varies with the intensity of magnetization. Thus, representing the constants of the two groups by K_{1A} and K_{1B} respectively gives

$$K_{1A}=\frac{K_1I_1}{I},\quad K_{1B}=\frac{K_1I_2}{I} \qquad (8)$$

The total crystal energy will be the sum of the respective field and anisotropy energies of the two domain groups. This reduces to the form

$$\begin{aligned}E=K_0+\frac{K_1I_1}{8I}(1-\cos 4\theta_1)-\\ H_{Ro}I_1(\cos\theta_1+\sin\theta_1)+\\ \frac{K_1I_2}{8I}(1-\cos 4\theta_2)-\\ H_{Ro}I_2(\cos\theta_2+\sin\theta_2) \qquad (9)\end{aligned}$$

If this equation is differentiated with respect to the variables θ_1, θ_2, and I_1, and if the assumption is made that the total saturation intensity of magnetization is represented by the sum of the components I_1 and I_2, we are left with the relation

$$\frac{K_1}{2}\sin 4\theta_1-H_{Ro}I(\cos\theta_1-\sin\theta_1)=0 \qquad (10)$$

and

$$\theta_1=\theta_2$$

The distribution of the domains can be calculated from equation 1, if it is written in the form

$$\begin{aligned}H_{Ro}=H_{Ta}-\\ N[I_1(\sin\theta_1-\cos\theta_1)+I\cos\theta_1] \qquad (1C)\end{aligned}$$

This provides the means of calculating the expected intensity of magnetization in the circumferential and longitudinal directions for various values of the applied magnetizing fields. If this computation is carried out, the calculated curves are again found to lie below the measured values in all cases.

For each of these calculated values, the constants I, N, and K, mentioned in the foregoing have been used. I, being a directly measured quantity, should be open to little question. The improper choice of the anisotropy constant will basically have an effect upon the slope of the curve rather than on the absolute value. The demagnetizing factor, however, plays an important part in setting the level of the calculated curves. It is usually thought of as a constant dependent only upon the geometry of the magnetic sample. This is strictly true only for a spheroidal specimen. In the case of a rod or a tube, the surface charge distribution will be to some extent dependent upon the intensity of magnetization, so that the demagnetizing factor will change slightly as the magnetization varies. To determine how N varies for the particular tube geometry used, a series of magnetization characteristics were taken in the axial or transverse direction for various values of H_{Ro} applied circumferentially. The curves show a definite trend toward a slope which decreases as H_{Ro} is increased. By computation from each of these curves, an effective value of demagnetizing factor shows that N varies with the circumferential field, approximately as shown in Fig. 8.

A more correct calculation of the magnetization characteristics should be ob-

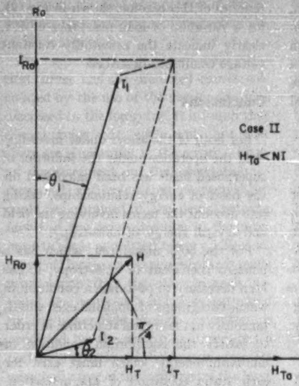
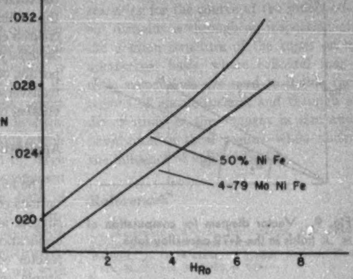


Fig. 7 (left). Vector relationships for superposed fields with H_{Ro} large

Fig. 8 (right). Variation of effective demagnetizing factor with circumferential field



tained, therefore, if the value of N is not held constant in the solution of equation 1(C), but is permitted to vary with H_{Ro} according to the curve of Fig. 8. Such a calculation has been carried out for the values of transverse field which correspond to the measured curves in the upper portion ($0 \leq H_{To} \leq 26.7$) of Fig. 6. The calculated curves are plotted in Fig. 6 for direct comparison.

A similar computation for the lower curves of Fig. 6 indicates that the effect of the variation of N is not appreciable here because of the large transverse applied field.

Some of the discrepancies between the measured and calculated curves may still be accounted for by an incorrect evaluation of the demagnetizing factor, but certainly not all may be. The points of least agreement are at low values of H_{Ro}, which is the region in which the approximation of ideal crystalline characteristics are poor. As is shown by the curvature of the magnetization characteristics for the tube with zero transverse field applied, a certain proportion of the magnetization changes are rotational, and the crystal orientation cannot be taken as perfect.

The greatest question at present, however, is the source of the seemingly independent domains which are easily influenced by the circumferential field even in the presence of extremely high axial fields.

4-79 Molybdenum Permalloy

The permalloy tube exhibits a considerably less rectangular hysteresis characteristic than the 50% nickel-iron alloy. This is undoubtedly due to the lack of orientation of the crystalline structure and the consequent need to rotate the magnetic domains against the forces of crystal anisotropy. Though not zero, the anisotropy constant of the permalloy is quite small. Bozorth and Walker[1] place the composition requiring zero anisotropy energy at about 76% nickel for the binary nickel-iron alloys which have an unoriented structure. On this basis, the 79% nickel alloy should have a small negative value, probably not exceeding -2×10^3 ergs/cm^3. The results obtained with the addition of molybdenum do not appear to differ with these findings. The magnetic characteristics of the permalloy tube are shown in Fig. 2.

Cross Fields

The general process of magnetization has been described in terms of two phenomena: the growth of domains by the motion of the boundary wall separating regions of different magnetic orientation within the crystal, and the rotation of whole domains away from the easy crystalline directions of magnetization toward a direction parallel to the applied field. In respect to the permalloy tapes, the field required to bring about boundary wall motion, roughly equal to the coercive force of the material, is extremely small. Further, because of the little energy of anisotropy required by the crystal, the magnetization vector can easily be rotated away from the preferred direction of magnetization, which in this case is parallel to the (111) axis of the individual crystals and is randomly oriented with respect to the tape dimensions. Therefore, the total magnetization vector should be aligned with the direction of the total applied field for values of field strength greater than one or two oersteds. If the assumption is made that this must be done, the characteristics of the tubular core may be calculated.

Referring to Fig. 9, if magnetizing fields are applied simultaneously in the rolling direction and the transverse direction, a resultant total magnetizing force H will lie at an angle θ_3 with respect to the transverse axis. This field is the sum of the H_{Ro} and H_T vectors. Thus

$$\tan \theta_3 = \frac{H_{Ro}}{H_T} \tag{11}$$

The effective field in the transverse direction, H_T, is dependent upon the transverse intensity of magnetization and the demagnetizing factor. Thus the H_T may be expressed as before in terms of the applied field and the component of intensity of magnetization in the transverse direction.

$$\tan \theta_3 = \frac{H_{Ro}}{H_{To} - NI \cos \theta_2} \tag{12}$$

where θ_2 is the angle between the total magnetization vector and the T axis.

In accordance with the assumption that the magnetic intensity vector and the intensity of magnetization vector are co-directed, equation 12 becomes

$$H_{To} \tan \theta_3 - NI \sin \theta_3 - H_{Ro} = 0 \tag{13}$$

This equation can be solved graphically for θ_3, permitting I_{Ro} to be computed for various values of H_{Ro} and H_{To} corresponding to the measured curves of Fig. 10.

The curves obtained from this computation have been plotted in Fig. 10. For each of these calculated points, the demagnetizing factor has been adjusted according to the data presented in Fig. 8. It will be noted that the agreement between the measured and calculated curves is quite good for small values of transverse field; but that as the values of H_{To} are made larger, an increasing discrepancy appears. This discrepancy is particularly large at low values of the circumferential field.

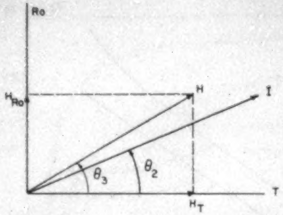


Fig. 9. Vector diagram for computation of fields in the 4-79 permalloy tube

At very small values of H_{Ta}, the calculated curves show a sharper rectangularity than do the measured values. This is undoubtedly accounted for by the error introduced with the assumption of a negligible anisotropy energy term. Since some anisotropy energy is required to achieve complete alignment of the magnetization and field vectors, it is not surprising to find the measured curves exhibiting a curvature rather than a sharp break when approaching saturation.

At high values of transverse applied field, the measured rolling direction magnetization is everywhere greater than the corresponding calculated points. The measured curves show a sharp rise in the circumferential intensity of magnetization at small values of applied field, indicating that the material is not acting exactly as has been assumed. In a manner similar to the 50% nickel-iron tape, the permalloy measurements show that a group of domains in the rolling direction are affected by a small circumferential field, even in the presence of a large transverse magnetizing force. It is concluded that the magnetization process cannot be completely described in terms of a rotation of the total I vector. For low values of H_{Ro}, the rolling direction flux density is increased either by domain growth through boundary wall motion, or by rotation of a portion of the magnetic domains out of the transverse direction through a substantial number of degrees.

Applications

The hysteresis loops shown in Fig. 4 for various conditions of cross magnetization suggest an interesting possibility for the application of transverse fields in magnetic reactor control problems. The standard simple series, or parallel power reactor circuits provide a control of alternating current in a load circuit which is nearly proportional to a d-c control current flowing in a control winding on the reactor.[3] This is less than satisfactory if the load being controlled has a varying impedance, since the output characteristic of such a reactor is essentially high impedance and provides a constant current drive. If, however, a single core reactor were provided with a means for the application of a variable transverse field, it should be possible effectively to control the saturation flux density of the reactor, and thus control the volt-time integral which the core would accept before firing. This in turn would make the firing angle, rather than the output current, a direct function of the control field, and make the output impedance characteristics relatively low. This is to say that the average voltage measured across a varying load impedance placed in series with such a reactor would remain constant.

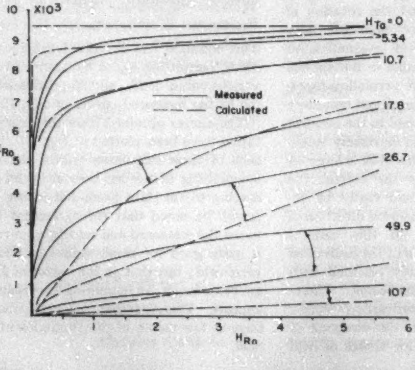
Such a reactor was built with a standard toroidal core of oriented 50% nickel-iron tape. A transverse control field was applied by placing the core between the poles of an electromagnet. The control characteristics of this reactor, shown in Fig. 11, for a variation of load resistance of 5:1, clearly indicate the essentially constant voltage conditions described.

Conclusions

For both of the alloys under investigation, the operation under the influence of superposed fields has been calculated on the basis of energy relationships, taking into account the terms involving the field and crystal energy of anisotropy.

For the 50% nickel-iron, which has a nonzero coefficient of anisotropy, it has been necessary to postulate a condition in which two groups of domains exist simultaneously in the crystal structure, in order to satisfy the equilibrium condition of minimum energy which must exist for each static condition of magnetization. These two domain groups differ in size and orientation in accordance with the relative magnitudes of the two superposed fields. This is unlike the situation which exists when a high-permeability material is magnetized by a single field, because as two domain groups change both size and orientation, the processes of domain rotation and boundary wall displacement occur simultaneously within the crystal. Under the influence of a single field, the orientation of the magnetic domains normally occurs first through domain growth in the crystal direction which is most nearly oriented with the field, and further increases in magnetization occur exclusively through domain rotation.

It was possible to make the assumption that the anisotropy constant was negligible for the 4-79 permalloy tape specimen, and that the field and magnetization vectors are at all times codirected. In this case, the two domain groups mentioned merge into a single



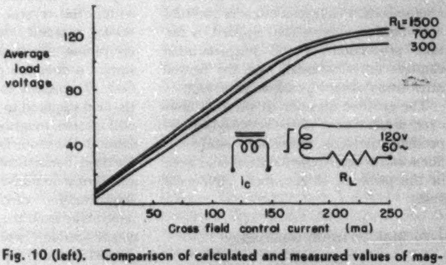


Fig. 10 (left). Comparison of calculated and measured values of magnetization in the 4-79 permalloy tube

Fig. 11 (above). Control characteristics of the cross-field reactor

group, and the field energy term alone determines the equilibrium of magnetization in the material.

In comparing the measured magnetization curves and the predicted curves calculated by the use of the energy relations discussed in the foregoing, it is found that discrepancies exist each time, but that there are similarities in the errors. First, an improvement in the correlation of measured and calculated characteristics could be obtained if the demagnetizing factor of the axial direction in the magnetic tubes is permitted to vary. Second, if one of the two superposed fields is made much larger than the other in each material, the material is not truly saturated by the larger field though saturation is predicted by the energy relations, but a certain proportion of the domains remains essentially free of the influence of the large field and is easily acted upon by the perpendicularly applied field.

In addition to eliminating the question of a demagnetizing factor from the general interpretation of the results, finding a means for closing the magnetic paths in both directions of measurement might permit a close inspection to be made of the source of the independent domains which do not appear to be influenced by large applied fields. Since this phenomenon has been found to be present in both of the alloys under discussion, it is apparently not primarily associated with the crystal anisotropy or magnetostriction properties of the materials, nor with the presence or lack of crystalline orientation. These factors vary greatly among the alloys studied, and yet the phenomenon is exhibited approximately equally by both. It is possible that the most profitable direction of study in searching for the source of the independent domains is through an inspection of the domain structure of the tapes on a microscopic basis, where colloidal magnetic suspensions are used to define the regions of magnetization, and through a direct study of the changes in size and orientation of these regions when under the influence of applied fields.

References

1. Magnetic Crystal Anisotropy and Magnetostriction of Iron-Nickel Alloys, R. M. Bozorth, J. G. Walker. *Physical Review*, New York, N. Y., vol. 89, 1953, p. 624.

2. Ferromagnetic Anisotropy in Nickel-Iron-Cobalt Crystals at Various Temperatures, L. M. McKeehan. *Ibid.*, vol. 51, 1937, p. 136.

3. Series-Connected Saturable Reactor with Control Source of Comparatively Low Impedance, H. F. Storm. *AIEE Transactions*, vol. 69, pt. 2, 1950, pp. 756–65.

Numerical Integration of Differential Equations on the 704 EDPM

H. H. ANDERSON
NONMEMBER AIEE

J. R. JOHNSON
NONMEMBER AIEE

Synopsis: To provide a convenient means for solving complex and nonlinear ordinary differential equations, a system has been devised which causes a *704* EDPM to behave like an analogue differential analyzer. Use of this system involves the interconnection of appropriate subroutines in a manner analogous to the interconnection of operational amplifiers in an analogue machine. The subroutines are designed so that the user may compile the required connections automatically, without much knowledge of digital computer coding. Solving differential equations by this method takes no more setup time than that required by analogue machines and in many cases may be a good deal faster. Integration is done by using Milne's formulas. The system automatically adjusts the integration interval to take advantage of portions of the solution which can be computed with large intervals at no sacrifice of accuracy. Nonlinear functions may be expressed by a table of values in storage.

A RELATED group of subroutines, called NIDA (Numerically Integrating Differential Analyzer) has been prepared for the *704* EDPM which facilitates the numerical solution of ordinary differential equations. By using the NIDA system, a person without much experience with *704* coding can quickly and conveniently solve a set of simultaneous ordinary differential equations. The system is applicable to equations of any order and degree, either linear or nonlinear.

The availability of these subroutines in conjunction with a *704* materially increases the value of the *704* to an engineering development organization. Long periods devoted to planning, coding, and debugging are eliminated and therefore the small problems which arise frequently can be economically solved by this method, often in a matter of hours. The NIDA system, on a *704*, can in most cases completely obviate the need for an analogue differential analyzer or similar devices. Differential equations which in the past have been solved only crudely, or perhaps not at all, by analogue methods, can now be numerically integrated at no sacrifice of time.

The operations which are performed by the *704* in executing the NIDA routines are essentially those used in desk calculator computations or other digital computer routines written to solve specific differential equations. By proper selection of conventions and linkages it is possible to place the necessary elements of differential equation integration routines in a few general-purpose subroutines. These subroutines can then be used as building blocks to form the complete program for any set of equations. The factor which was held most important in the design of NIDA was convenience of use. The parts were written so that compiling for a problem could be done easily and accurately by one not skilled in the art of *704* coding. Secondary in importance is the speed at which integration progresses in a solution. Fortunately, the *704* has enough inherent speed and flexibility so that very little speed was sacrificed.

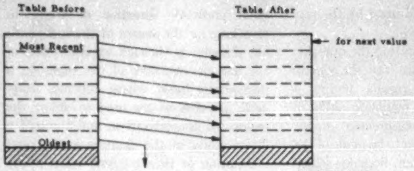
Integration

Integration is accomplished by use of Milne's prediction and correction formulas. The information required by the integration subroutine consists of two tables of values, one table for present and past values of the integrand and one table for past values of the integral. Each table is composed of seven entries, each entry having been obtained from previous computation. As the solution progresses, the entries in the tables are shifted, dropping the oldest, and making room for the next value to be computed.

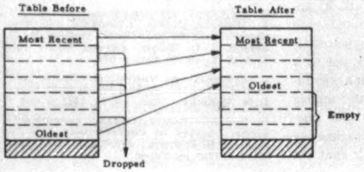
Paper 56-697, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 30, 1956; made available for printing May 29, 1956.

H. H. Anderson and J. R. Johnson are with the International Business Machines Corporation, Poughkeepsie, N. Y.

The authors are indebted to Dr. Gene M. Amdahl who is responsible for the conception of a large part of NIDA and who freely gave his assistance in the actual preparation of the subroutines



(a)
Normal Up-dating of Table



(b)
Interval Doubling of Table

Fig. 1. Variable table handling

A table, as described is required for every variable (each derivative is considered to be a separate variable) involved in the equation. The process of integration thus consists of entering a subroutine, specifying the location in storage of the two tables to be used by a subroutine.

For a given increment of the independent variable, that is, the variable of integration, each derivative is operated on twice. First, a formula makes use of previous values of the function and predicts a value of the integral for the present This predicted value is stored in a temporary location at the end of the table. Subsequently, a correction formula is applied to the table, including the predicted value, and a corrected value is obtained. This value is stored at the head of the table, representing the most recent value of the integral.

After having computed the corrected value, a comparison is made between the results of the two computations and if the two values do not agree within a specified amount, a process is initiated which effectively reduces the size of the integration interval being used. If the results of the prediction and correction computations agree too closely, a process is initiated which causes the integration interval to be doubled. This automatic adjustment of the interval size goes on without special arrangements on the part of user of the system, and affords a method of obtaining both small interval size when needed for accuracy as well as a large interval size for speed, when the accuracy requirements are satisfied.

A discussion of the mathematics of the integration subroutine is presented in the Appendix.

Algebra

The algebraic operations required in the solution of a differential equation, combined with integrations, completely specify the equations. Such operations as multiplication of derivatives by constants or other variables, summing of terms, and any other necessary arithmetic operations are coded directly in *704* machine language by the user of the NIDA system. This coding is straightforward, for the most part, and because of the built-in floating point arithmetic feature of the *704*, the coder need not concern himself with considerations of radix point or overflow. The use of *704* instructions for the algebraic portion of the equation is, in no way, a disadvantage of the system. On the contrary, it is this feature which most clearly establishes the superiority of NIDA over analogue methods. Multiplication, function generation, and other nonlinear operations (traditionally the weak points of analogue methods) are the most precise and easily performed functions in using NIDA. The user also has at his disposal all of the powerful logical operations available in the *704* instruction repertoire. The logical operations are particularly valuable when equations change form as the variables reach limits. An example of this might be the case of a 2-loop electric circuit, wherein the loops are coupled by an ideal diode. A logical testing instruction could be employed to sense when the diode current attempts to reverse and at this point switch the program to a different equivalent circuit.

Up-Dating

The essential core of the NIDA system, that which accounts for its ease of use and flexibility of application, resides in a subroutine called "up-dating." This is the part of the system which links all of the other parts together logically and exerts supervision and control. The update routines sees to it that the entries in the variable tables are shifted at the proper time. Up-date also keeps a running record of the value of the independent variable and the integration interval being used at any time.

As each correction integration is performed by the integration subroutine, a note is left with the up-date routine regarding whether the interval size should be increased, maintained, or decreased. The up-date routine, when executed, examines these notes regarding the accuracy of integration of each variable in the equation and if any one indicates that the interval should be decreased, then it is decreased. If all the notes indicate that the interval should be doubled, then it is doubled. In other cases the interval size is not disturbed.

When the solution of the differential equations begins each variable table contains only the initial value of the function, stored in the first cell of the table. Since Milne's formulas require four values of the variable, a special set of starting formulas is used to get started. Because it is convenient to have only one integration subroutine to link, the up-date subroutine works in conjunction with the integration subroutine, directing it to use the proper starting formula during the starting phase and also instructing it when to use a prediction or correction formula later on. If the situation arises which requires that the integration interval be reduced (error too large), the up-date routine causes the integration procedure to restart. That is, using the most recent computed value of each function as a new initial condition, the integration routine again employs the starting formulas with a very small starting interval. All values of each variable except the most recent are ignored.

If conditions arise which make it possible to increase the integration interval, this is done by doubling the interval and then discarding alternate entries in each variable table. This reduces the

Vol. 7
Pt. 1
1956

number of entries in each table from seven to four, and so the next time at which doubling can be performed is restricted to a time when seven entries have again accrued. The process of handling the variable tables is illustrated in Fig. 1.

Printing and Table Look-Up

Included in the NIDA system are two "convenience" subroutines, printing and table look-up.

The printing subroutine is capable of translating from floating binary and printing floating point decimal numbers, six to a line. The mode of entry to the print subroutine is flexible enough so that any number of quantities may be printed per entry.

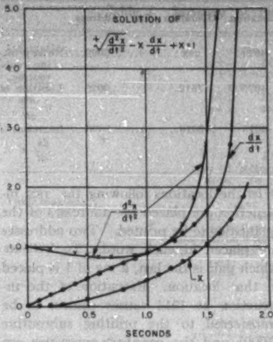
Frequently, the integration interval is less than the interval for which results are to be printed. However, the printing of results is limited to those points which are desired. This control of printing is exerted by the up-date subroutine.

In order to handle easily any complicated functions which appear in the equations being solved, a table look-up subroutine can be used with a stored table of functional values and can supply the main routine with values when required. The table look-up subroutine uses a polynomial of high order for interpolation so a relatively small number of tabular values needs to be stored to achieve accuracy comparable to integration.

Stability of the Integration Formulas

Original plans for the logical organization of NIDA called for a halving of the integration interval when the predicted and corrected values for any integration did not agree within the specified allowable error. This method of reducing the interval size proved to be unsatisfactory because of the instability of Milne's integration formulas.[1] To avoid stability difficulties it was decided to restart the process completely, with the latest computed values as initial conditions, when reduction of interval size is called for. It was found by experiment that repeated restarts did not significantly affect the over-all accuracy of the solution.

Restarting has the disadvantage that it consumes a relatively large amount of time, as compared to interval halving. If Milne's integration formulas are replaced by stable formulas, it may be possible to use the interval halving procedure with a consequent decrease of solution time in most problems.



t	$\ddot{x}$	$\dot{x}$	x
0.0	1.0000000	0.00000000	0.00000000
0.1	0.99103121	0.099692323	0.0049921786
0.2	0.96835172	0.19774931	0.019883268
0.3	0.93814277	0.29311245	0.044451625
0.4	0.90593902	0.38531221	0.078400309
0.5	0.87644380	0.47439127	0.12141004
0.6	0.85366361	0.56082543	0.17319030
0.7	0.84127777	0.64547088	0.23351550
0.8	0.84319131	0.72955768	0.30226595
0.9	0.86435011	0.81475000	0.37946389
1.0	0.91203908	0.90330991	0.46532810
1.1	0.99824823	0.99843620	0.56034474
1.2	1.1445395	1.1049475	0.66539494
1.3	1.3933872	1.2307208	0.78197631
1.4	1.8386521	1.3900359	0.91265715
1.5	2.7245658	1.6125893	1.0620914
1.6	4.8728514	1.9739096	1.2398006
1.7	12.710409	2.7444025	1.4705124
1.8	146.83713	6.9352876	1.8731421

Fig. 2. Solution of the example differential equation

An Example of How to Solve an Equation Using NIDA

The foregoing has been an attempt to explain in general terms what the NIDA system is and how it works. In what follows, it will be shown in detail how the system is applied to the solution of an example differential equation (see Fig. 2). Consider

$$+\sqrt{\frac{d^2x}{dt^2}}-x\frac{dx}{dt}+x=A \tag{1}$$

The first step in preparing this equation for NIDA solution is to rearrange it, solving for the highest order derivative. This is commonly done in preparing equations for solution on most differential analyzers, and results in

$$\frac{d^2x}{dt^2}=\left(x\frac{dx}{dt}-x+A\right)^2 \tag{2}$$

An examination of this equation shows that three variables appear which must be stored in tabular form for the application of Milne's integration formulas. Accordingly, storage assignments in the *704* core storage must be made for the tables d^2x/dt^2, dx/dt, and x. As mentioned before, these tables will occupy eight words each, and the variable expressed by each table will be addressed by the location of the first word of the able. Suppose that the table locations are selected as follows:

x:3000 (3000 through 3007) (In what follows, all addresses and locations are expressed in the octal system.)

$\frac{dx}{dt}$:3010 (3010 through 3017)

$\frac{d^2x}{dt^2}$:3020 (3020 through 3027)

and assume the constant A is stored at location 3030. A *704* instruction routine can now be written which will accomplish the algebra implied by equation 2. Reference to a *704* "Manual of Operation" will clear up any question regarding the function of the following instructions. Suppose the instructions of this routine are stored beginning at location 4000, as follows:

Location	Operation	Address	Result
4000	LDQ	3010	Get $\frac{dx}{dt}$ into multiplier register.
4001	FMP	3000	Form product $x\frac{dx}{dt}$ in accumulator.
4002	FSB	3000	Subtract x from accumulator.
4003	FAD	3030	Add A to accumulator.
So far, $x\frac{dx}{dt}-x+A$ has been computed.			
4004	STO	3031	A temporary storage location.
4005	LDQ	3031	Get sum in multiplier register.
4006	FMP	3031	Form square of sum.
According to equation 2 this quantity is $\frac{d^2x}{dt^2}$, therefore			
4007	STO	3020	Store result in location of $\frac{d^2x}{dt^2}$

The routine stored at locations 4000–4007 expresses the differential equation which is being solved by expressing the highest order derivative in terms of the rest of the equation. The solution routine will be complete when the high-order derivative so computed is successively integrated to yield lower order derivatives. This is done by using the integration subroutine as shown in the following table.

Location	Operation	Tag	Address	
4010	TSX	(4)	5500	Enter the integration subroutine.
4011		3010	3020	Which variables to use.

The instruction at address 4010 causes the program to interrupt its normal sequence and to transfer control to the first instruction of the integration routine, at 5500. The location 4011 contains two addresses which the integration subroutine requires. The left address (in the "decrement" part of the word) 3010, is the location of integral function, dx/dt, and the right address 3020 is the location of the integrand function, d^2x/dt^2. The Milne formula will be applied to the proper numbers appearing in the two tables specified, and the value of the integral computed will be placed in the table at 3010. The number at 3010 is thus the "current" value of the function dx/dt. After completing this computation, control is transferred back to location 4012, and the next instruction is taken from there. The previous sequence must be repeated to obtain x from dx/dt, so

Location	Operation	Tag	Address	
4012	TSX	(4)	5500	Enter integration subroutine.
4013		3000	3010	Which variables to use.

This time the integrand is the previously computed dx/dt and the integral is x, which is placed in the table at 3000, becoming the "current" value of x. The instructions written so far complete the mathematical portion of the routine to solve equation 2.

For the process to begin and proceed properly, it is necessary to store at locations 3000 and 3010 the initial values of the variables x and dx/dt, respectively. The execution of the instructions 4000 to 4007 will compute the initial value of d^2x/dt^2 and the instructions 4010 to 4013 will predict the next values (at $t_0 + \Delta t = t_0 + h$) of x and dx/dt.

The next step in the program could be to print the values of the computed functions. A subroutine is provided for this purpose which has the ability to print an unlimited number of quantities, six to a line, in floating point decimal notation. It would be desirable to print the independent variable t, d^2x/dt^2, dx/dt and x.

Thus:

Location	Operation	Tag	Address	
4014	TSX	4	5000	Enter printing subroutine
4015	7612		3020	Location of t; $\frac{d^2x}{dt^2}$
4016	3010	1	3000	Location of $\frac{dx}{dt}$; x

In the locations following the TSX instruction are placed the addresses of the quantities to be printed. Two addresses are placed in each word. To indicate which pair is the last, a tag of 1 is placed in that location. Execution of the instruction at 4014 causes control to be transferred to the printing subroutine (5000) and the specified quantities are translated from floating binary to floating decimal and printed. Then control is returned to the location following the location of the last pair of addresses to be printed. Any number of pairs of addresses may follow the TSX at 4014, however, the last pair must always be identified by a tag of 1. The printing subroutine functions as described upon its first entry (that is, the first time the instruction at 4014 is executed) however, on subsequent entries, it may or may not print depending upon whether the solution has reached a "printing point." To avoid the printing of excess data, the printing subroutine will ignore all entries that are not coincident with a prescribed printing interval. The printing interval is a quantity that must be stored in location 7603 before starting the program. In this example problem, a suitable printing interval might be 0.1 second, therefore, 0.1 should be stored at location 7603.

The last step to complete the program for the solution of equation 2 is to cause the advance of the independent variable t to its next value, $t_0 + h$. This implies that all the variable tables must be shifted, as required by the fact that the integration routine is producing new values of each function for which space must be provided. These two operations, as well as others, are taken care of by the up-date routine, which is entered thus:

Location	Operation	Tag	Address	
4017	TSX	4	7200	Enter "up-date"
4020	3000		3010	Variable table addresses.
4021	3020	1	0000	

After the up-date entry TSX, the addresses of each variable table are given so that these tables can be shifted. The tag of 1 appears in the last word for reasons discussed, in connection with printing. Having completed all the operations required to solve the equation, the final instruction serves to return the program to the original instruction, at 4000, so that the process repeats itself:

Location	Operation	Tag	Address	
4022	TRA		4000	Return to beginning.

Consider the over-all action of the program stored from 4000 to 4022. When the execution begins, the initial conditions are used to compute the high-order derivative. Then the integration routine predicts the value of the lower order derivatives for $t = t_0 + h$. This is termed a prediction pass. At the end of a prediction pass, the up-date routine advances the independent variable to $t_0 + h$ and shifts the tables. The control then returns to 4000 and a corrected value of the high-order derivative is computed from the predicted values of the lower order derivatives. From this corrected high-order derivative, corrected values of the lower derivatives are "integrated." This is termed a correction pass. Throughout the solution of the equation correction passes are alternated with prediction passes, each pass, of course, using the same instruction loop. At the end of a correction pass, the up-date routine compares the results of the two passes and determines how closely the computed values agreed. For this purpose an initially stored value of allowable error (place at 7604) is used. If the predicted and corrected value for any integration did not agree within the amount of this error, the integration interval is restored to a very small, starting value. If the predicted and corrected values for all integrations were closer than 1/100 of this error, then the integration interval is doubled to allow greater computation speed. This is done automatically by the up-date subroutine and the necessary changes in the variable tables to effect the interval change are taken care of.

The magnitude of the allowable error to be used can be determined by consideration of accuracy required in the solution. If only three decimal digits are desired in the results, an allowable error of 10^{-3} or 10^{-4} should suffice. As the magnitude of the allowable error is reduced, the solution time required will, of course, increase because integration will proceed with a smaller interval in the independent variable.

The interval of integration can be doubled as often as is permissible, however its size will not be allowed to exceed the printing interval, so that no printing interval will be skipped.

Simultaneous Differential Equations

If several simultaneous differential equations are to be solved, the procedure to follow is similar to that used for the example. Each equation should be solved for its highest order derivative. If the high-order derivative appears in more than one equation, it may be possible to eliminate the derivative and reduce the number of equations. The ideal situation is obtained when the coupling between equations is by lower order terms only.

If this form can be obtained, then all the algebraic operations for all the equations can be done in the first portion of the solution loop, followed by integration of all the derivatives in each equation.

If some coupling between equations is by highest order derivative, then the equation having this derivative as its left member must be summed first in the algebra so that the derivative can be used to form other equations in the group.

Table Look-Up

Suppose that in the equation to be solved, a term such as e^x had appeared. The easiest way to handle this would be to make use of the table look-up subroutine provided in NIDA. At some point in the algebraic portion of the solution loop the value of this function would be required and could have been obtained thus:

Location	Operation	Tag	Address	
4005 (say)	TSX	4	6100	Enter table look-up.
4006	4100		3000	e table; x

The location following the TSX instruction contains two addresses required by look-up. The left address indicates where the table of values of the exponential function is stored and the right address is the location of the argument whose function is to be found in this case x. The look-up routine fits the table of values with a seventh order polynomial, and interpolates from this. When the control is returned to the main loop at 4007, the value of e^x appears in the accumulator and must be used or stored at that time. The machine will stop to signal an error if an attempt is made to extrapolate the stored table. The table of values of the function stored beginning at 4100 must have the following form:

Location

4100 : The tabular interval (floating binary).
4101 : The number of tabular entries in fixed binary with the binary point considered to follow position 35 of the word.
4102 : The value of the argument corresponding to the first entry in the table (floating binary).
4103 : First tabular entry (floating binary).
4104 : Second tabular entry, etc.

Table I. Summary of Nida Linkage Conventions (Octal)

Integration	q:TSX	(4)	5500
	$q+1$: $L(y)$		$L(y)$
	$q+2$:Return		
Printing	q:TSX	(4)	5000
	$q+1$: $L(a)$		$L(b)$
	$q+2$: $L(c)$		$L(d)$
	⋮ ⋮		⋮
	$q+i$: $L(r)$	(1)	$L(s)$
	$q+i+1$:Return		
	Note: To force printing on every iteration, replace address 5000 by 5015.		
Table look-up	q:TSX	(4)	6100
	$q+1$:L(table)		L(argument)
	$q+2$:Return		
	See section 6 for table format convention.		
Up-date	q:TSX	(4)	7200
	$q+1$: $L(x)$		$L(x)$
	$q+2$: $L(y)$		$L(y)$
	⋮ ⋮		⋮
	$q+i$: $L(r)$		$L(s)$
	$q+i+1$:Return		

Initial conditions for all variables are stored in the first table location.
Independent variable is stored at location 7612.
Permissible error is stored at location 7604.
Printing interval is stored at location 7603.
Store first instruction of main program at location 4000.
Start operation by transferring to location 7700.

Loading Information

The instructions comprising NIDA are contained in 63 binary cards and may be loaded with any absolute binary loader. NIDA occupies storage locations 5000 to 7777. The instructions of the solution loop may be punched and loaded in any way desired. All computations in the NIDA system are done in floating binary, consequently the data must be so loaded and the solution loop must use floating point arithmetic when dealing with these data.

The solution loop instructions must be arranged so that location 4000 is the first instruction to be executed. After loading all instructions and data, start by transferring to 7700.

The NIDA routines are also available in symbolic form.

Speed

The approximate execution times for the NIDA subroutine are as follows:

Integration = 3.66 milliseconds per integration per variable. This includes both prediction and correction
Up-dating = 2.74 milliseconds per interval (includes functions following both prediction and correction passes)
Table look-up = 7.5 milliseconds per use

Thus, in the foregoing example, the approximate time required per integration interval would be about:

2 integrations at 3.66	= 7.32
1 up-date at 2.74	= 5.48
Main loop instructions (approximate)	= 0.7
	Total 13.5 milliseconds per point

Printing is done at 150 lines per minute, interspersed with the computation.

A summary of NIDA linkage conventions is given in Table I.

Summary

This paper has described a convenient and time-saving method of integrating differential equations using the 704 EDPM. The method is applicable to all ordinary equations of arbitrary order, linear or not. An example was given which it is hoped was sufficiently explicit to instruct interested persons in the use of the system.

Appendix. Milne's Method for Solving Differential Equations

The method of Milne selected for solving a differential equation requires that four starting values of the function sought are known. How to obtain these starting values will be discussed later. To proceed with the explanation, let it be assumed that the problem is to find a particular solution of

$$y' = f(x,y) \tag{3}$$

say, the solution passing through (x_0, y_0). Let it be further assumed that one has found the values of y for $x = x_0 + h$, $x_0 + 2h$, and

x_0+3h. Milne's method for extending the solution consists in finding y for $x=x_0+4h$ by the equation (predictor)

$$y_4=y_0+\frac{4h}{3}(2y_1'-y_2'+2y_3') \qquad (4)$$

where

$$y_i=y(x_0+ih)$$

and

$$y_i'=f(x_0+ih,\ y_i)$$

$$i=1,\ 2,\ 3,\ 4.$$

The values of y_1', y_2', and y_3' are computed by substituting (x_1, y_1), (x_2, y_2) and (x_3, y_3) into equation 3. Next one obtains y_4 by equation 4 and then y_4' by substituting (x_4, y_4) into equation 3. The values of y_4 and y_4' thus found are tentative values of y and y' corresponding to $x=x_4=x_0+4h$. These values are then checked by the equation (corrector)

$$y_4=y_2+\frac{h}{3}(y_4'+4y_3'+y_2) \qquad (5)$$

Equations 4 and 5 obtain from the following relation

$$\int_{y_i}^{y_n} dy=\int_{x_i}^{x_n} y'dx$$

or

$$y_n=y_i+\int_{x_i}^{x_n} y'dx \qquad (6)$$

by replacing the integral with an approximate integral. In particular the approximation is achieved by assuming that y' is a polynomial. The error in equation 5 may be shown to be

$$\frac{-h^5}{90} f^{(4)}_{(u)} \qquad (7)$$

where u is some number between x_0 and x_0+4h. The error in equation 4 is

$$\frac{28}{90} h^5 f^{(4)}_{(v)} \qquad (8)$$

where v lies between x_0 and x_0+4h.

In the application of Milne's method, the exact values of the error of equations 7 and 8 cannot be computed since, in general, the function of x representing y' is not known. Further, the derivations of these error equations fail to stipulate the precise values of u and v. When $f(x)=y'$, the error equations 7 and 8 become

$$\frac{-h^5}{90}\times\frac{d^5y}{dx^5}$$

and

$$+\frac{28h^5}{90}\times\frac{d^5y}{dx^5}$$

respectively. The fifth derivative under ordinary circumstances remains essentially constant over the interval from x_0 to x_0+4h. Accordingly, the difference between the values for y_4, as given by

$$y_{4p}=y_0+\frac{4h}{3}(2y_1'-y_2'+2y_3') \qquad (9)$$

and

$$y_{4c}=y_2+\frac{h}{3}(y_2'+4y_3'+y_4') \qquad (10)$$

is

$$\frac{29h^5}{90},\ \frac{d^5y}{dx^5}$$

One twenty-ninth of this is the error in y_4 as given by equation 7. It is clear that the ratio of error may be defined as

$$\frac{2}{29}\left[\frac{|y_{4p}-y_{4c}|}{|y_{4p}|+|y_{4c}|}\right]$$

Experience dictates that this is too optimistic and that a more accurate result obtains by replacing 29 with the divisor 10. Thus one may define

$$\text{Ratio of error}=E=0.2\,\frac{|y_p-y_c|}{|y_p|+|y_c|} \qquad (11)$$

If equation 11 is significant, Milne's method may be continued with a smaller value of h. For example, if one halves the value of h, calling the last four established y-values y_0, y_1, y_2, y_3, one needs the values designated by $y_{3/2}, y_{5/2}$, viz., the values corresponding to

$$x=x_0+\frac{3h}{2}$$

and

$$x=x_0+\frac{5h}{2}$$

By applying equation 6

$$y_{5/2}=y_2+\int_{x_2}^{x_2+\frac{h}{2}} y'dx$$

and

$$y_{3/2}=y_1+\int_{x_1}^{x_1+\frac{h}{2}} y'dx$$

Replacing y' by a polynomial and performing the integration one obtains

$$y_{5/2}=y_2+\frac{h}{384}(7y'_{-1}-37y'_0+197y'_1+25y'_2) \qquad (12)$$

As mentioned previously, the Milne method requires the possession of four consecutive points on the curve before it can be started. In the same manner that the Milne method finds a fifth point on a curve, given four points, one can find a second point given one, a third point, given two, and a fourth point, given three. This is accomplished by applying equation 6. y' is replaced by a polynomial. The order of the polynomial increases as one advances along the curve. The starting formulas obtained in this manner are as follows, with the estimated error given in the right-hand column:

Integration A $y_{1p}=y_0+hy_0' \qquad \frac{h^2}{2}y''$ (13)

Integration B $y_{1c}=y_0+\frac{h}{2}(y_0'+y_1')-\frac{h^3}{12}y'''$ (14)

Integration C $y_{2p}=y_0+2hy_1' \qquad \frac{h^3}{3}y'''$ (15)

Integration D $y_{2c}=y_0+\frac{h}{3}(y_0'+4y_1'+y_2')-\frac{h^5}{90}y^{(5)}$ (16)

Integration E $y_{3p}=y_0+\frac{3h}{4}(y_0'+3y_2') \qquad \frac{3h^4}{8}y^{(4)}$ (17)

Integration F $y_{3c}=y_0+\frac{3h}{8}(y_0'+3y_1'+3y_2'+y_3')-\frac{3h^5}{80}y^{(5)}$ (18)

Equations 4 and 5 of Milne's for extending the solution may be referred to as integration G and integration H, respectively.

Reference

1. Quarterly of Applied Mathematics, H. J. Gray. Vol. 12, p. 133.

A Field Survey of Transformer Oil Quality

W. J. DEGNAN
MEMBER AIEE

E. J. SHIMANSKI
NONMEMBER AIEE

REDUCTION of power transformer insulation by the recently recommended two full basic insulation levels[1] has placed renewed emphasis on oil quality. The design of corona-free transformers has eliminated the use of sharp-edged electrodes in highly stressed locations, favoring the more uniform dielectric field conditions. Because contamination in these uniform fields has a much greater influence on dielectric strength that in nonuniform fields, the question of oil quality is of further importance.

Not only must the transformer oil be of high quality during factory tests but it is essential that it does not degrade under service conditions or, if degradation eventually exists, some means must be available to determine the point at which preventive maintenance steps must be taken. If transformer oil meets these requirements, then the full cost-saving advantage of the reduced insulation can be realized.

To determine the quality of oil in transformers that have been in service for long periods of time, it was decided to conduct a nation-wide survey of many such transformers. The regions of the country and types of transformers selected were such that the results obtained are representative of large power transformers located anywhere in this country or in similar climatic regions. The regions correspond to the eight areas indicated numerically on the map shown in Fig. 1. The units sampled were all General Electric Company power transformers in actual service for periods up to 17 years. The transformers were generally well-loaded units rather than those "idling" on the line and were rated 115 kv and higher with a minimum of 10,000 kva. A variation in oil-preservation methods was obtained with three of the basic types being utilized in present-day design of power transformers (conservator, gas-oil-sealed, and automatic gas-sealed). Typical modes of cooling were also represented in the units (self-cooled, forced-air-cooled, forced-oil-cooled, and a few water-cooled). Within these limits, the units were selected at random. All sampling was conducted during the fall and winter of 1955.

Sampling Technique

A preliminary study was conducted locally as a pilot operation to provide information on techniques to be utilized in the broader investigation, one outcome of which was the development of a reliable outdoor sampling technique for large transformers.

Of primary concern were the containers into which the samples were to be drawn. It was found necessary to provide clean, new-oil-filled containers from the central testing laboratory (Pittsfield) to insure uniformity and eliminate the possible influence caused by a doubtful sample bottle condition. Each unit was sampled from the bottom sample valve into two 1-quart amber-glass bottles provided with a foil-lined cap. The clearest picture of the sampling technique can be given by listing the procedure as outlined in the instructions to the field representatives:

1. Do not sample on very cold days or during rain, snow, or fog.
2. Sample only from units with a bottom valve and sampling valve. Do not use the main opening in the valve.
3. Wipe the sampling valve thoroughly with a clean, dry cloth.
4. Bleed off 1 quart of oil into a separate container and discard.
5. Empty two of the 1-quart bottles which were shipped oil-filled and fill each half full of oil directly from the sample valve; shake and discard. Fill half full again, shake and discard.
6. Fill each of the two bottles to the height of the neck and allow oil to impinge on the wall of the bottle. Cap tightly with the foil-lined screw cap provided.
7. Sample units with up to 20 years of service.

Field data were also accumulated to determine the history of the unit and its oil with particular emphasis on any filtration or reconditioning of the oil.

Oil Tests and What They Signify About Oil

A very concise description of the various electrical, physical, and chemical tests usually employed in evaluating insulating oils has been set up in the unpublished report by the Subcommittee on Insulating Fluids of the AIEE Committee on Transformers. This guide quite adequately describes the tests and, therefore, a verbatim inclusion of their descriptions is made in this paper. Added to the discussion are comments regarding the ASTM (American Society for Testing Materials) dielectric strength tests and the entire sections on the water content test and the use of the VDE (Verein Deutscher Elektrotechniker) test electrodes. All tests in this survey conformed to the following outlined procedure. Reference to the ASTM number will provide a complete description of each test.

DIELECTRIC STRENGTH

ASTM Test (ASTM D-877)

"The dielectric strength of an insulating oil is the average voltage gradient at which electrical failure or breakdown occurs under prescribed conditions. This test is made to determine if an oil has adequate electrical strength and is not necessarily an indication of its other properties."

A low dielectric strength is indicative of the presence of contaminating agents such as water, dirt, or conducting particles. The true significance of a dielectric strength value depends upon how closely it is a reflection of conditions within the transformer being tested as well as upon the actual stress that will be applied to the oil in the unit. Generally a dielectric strength of over 22 kv is considered acceptable for used oils.

The VDE Test Electrodes

The testing of oils for dielectric strength by means of more uniform field electrodes, and particularly the spherical surfaced VDE type, is more specifically the topic of a paper[2] which describes another portion of the broad program dealing with oil serviceability. That work shows the more uniform field gap to be more sensitive to those factors influencing oil strength and, therefore, a better criterion than the previously used nonuniform electrodes

---

Paper 56-805, recommended by the AIEE Dielectrics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 27, 1956; made available for printing June 22, 1956.

W. J. DEGNAN and E. J. SHIMANSKI are with the General Electric Company, Pittsfield, Mass.

The authors wish to express their appreciation to P. E. Flanagan, J. J. Frank, Mrs. V. Judware, and S. W. Kernaghan for their participation in this survey.

Table I. Field Oil Electrical, Chemical, and Physical Properties

Sample No.	Oil Age	Cooling Method*	Preservation Method†	Water Content, Parts per Million‡	Dielectric Strength, Average Kv-Rms		Power Factor, Per Cent		Resistivity, 10^{12} Ohms-Centimeter		Acidity, Milligrams KOH per Gram	ASTM Color	Viscosity, Seconds	Interfacial Tension, Dynes per Centimeter
					VDE	ASTM	25 Degrees Centigrade	100 Degrees Centigrade	25 Degrees Centigrade	100 Degrees Centigrade				
Region 1—Pacific South Coast														
1	5	OA	GS	11	57.2	41.4	0.06	2.7	19.5	0.3	0.006	1½	54.0	33.4
2	7	FA	GS	7	57.4	43.4	0.08	2.8	20.0	0.2	0.019	2	55.0	31.5
3	8	OA	GS	10	52.2	39.2	0.08	2.0	26.0	1.7	0.025	1½	57.2	28.4
4	17	OA	GS	9	58.4	36	0.05	1.8	29.3	1.3	0.025	1½	54.0	24.7
5	1	FA	GS	9	57.6	43	0.09	4.2	13.0	0.1	0.006	1	54.0	30.1
6	16	OA	GS	8	57.8	42.6	0.1	2.3	12.4	0.65	0.051	2½	54.2	21.0
7	2	FA	GS	16	56.2	38.6	0.07	3.8	15.0	0.12	0.006	1½	54.0	32.1
8	6	FOA	GS	12	56.8	41	0.08	3.1	18.0	0.2	0.013	2½	54.3	31.2
9	12	FA	GS	18	59.4	40.8	0.06	1.6	22.0	1.5	0.025	1	55.0	22.0
10	16	OA	GS	11	52.6	36.6	0.03	4.3	13.0	0.5	0.045	2	54.2	28.1
11	7	FA	GS	9	55.8	41.6	0.1	4.4	10.3	0.65	0.025	2	54.6	28.3
Region 2—Gulf Coast														
12	5	FA	C	7	54.8	43.8	0.07	1.5	26.0	1.68	0.013	1½	55.6	33.3
13	6	FA	C	7	59.0	43.2	0.08	3.4	13.0	0.91	0.013	1½	54.2	30.1
14	6	FA	C	7	58.4	42.4	0.07	5.2	9.75	0.65	0.019	2	54.0	29.3
15	6	FA	C	8	60.0	45.2	0.06	2.2	19.5	1.17	0.019	1½	54.2	29.0
16	6	FA	C	15	58.0	44.6	0.08	2.2	13.0	0.86	0.025	2	55.0	26.8
17	2	FA	GS	8	56.8	41.2	0.05	1.0	39.0	2.1	0.006	1	55.0	32.2
18	8	OA	GS	17	42.4	35.8	0.04	0.8	67.2	3.3	0.045	1½	55.4	26.2
19	2	FA	GS	12	55.8	29.4	0.1	1.1	39.0	3.3	0.006	1	55.0	36.6
20	5	FA	GS	8	52.6	37.2	0.06	2.1	13.0	1.5	0.013	1¼	55.4	30.0
21	4	FOA	GS	8	55.6	45.4	0.07	1.8	25.0	1.56	0.025	1½	55.0	23.6
Region 3—South East														
22	2	FOA	C	9	42.8	38.0	0.09	3.7	13.0	0.8	0.006	2	54.2	27.8
23	2	FOA	C	12	36.8	36.0	0.1	3.5	11.8	0.9	0.013	2	54.0	26.5
24	4	FOA	GS	12	45.0	40.4	0.13	2.3	24.5	1.3	0.032	2	55.0	25.5
25	2	FOA	GS	14	52.4	38.4	0.06	1.3	45.5	2.0	0.013	1½	55.2	25.7
26	10	FOW	GS	12	51.4	40.6	0.04	2.1	15.8	1.3	0.036	2	54.3	22.4
27	5	OA	GS	16	50.0	35.8	0.05	1.0	65.0	3.9	0.019	1	54.4	30.5
28	5	FA	GS	13	48.6	40.6	0.06	1.4	39.0	1.95	0.019	1½	54.0	29.7
29	2	FA	GS	15	46.6	39.6	0.07	0.6	65.0	6.5	0.006	1	54.0	36.9
30	5	FA	GOS	10	50.8	43.0	0.07	3.8	13.0	0.98	0.013	2	54.4	32.0
Region 4—Pacific North Coast														
31	4	FA	GOS	11	47.4	39.2	0.08	1.8	39.0	2.6	0.013	1	54.8	32.2
32	4	FA	GOS	8	51.2	38.2	0.09	2.1	19.5	2.1	0.019	1	54.2	35.9
33	2	FA	GS	7	51.4	36.8	0.05	0.6	97.5	7.15	0.006	1/2	54.4	35.1
34	12	FA	GS	9	55.6	30.8	0.07	2.7	32.5	0.86	0.006	1	55.0	31.6
Region 5—Mid-Atlantic Coast														
35	1	FOA	GS	12	47.8	36.0	0.06	0.6	83.9	13.0	0.019	1½	54.6	32.0
36	5	FA	C	9	51.2	38.8	0.3	5.0	7.8	0.8	0.019	1½	54.8	28.8
37	5	FOA	C	20	45.6	38.8	0.2	6.0	10.0	0.8	0.013	1½	54.0	32.0
38	3	FOA	C	13	50.2	41.6	0.08	1.7	13.2	2.0	0.013	1½	55.8	31.9
39	4	FOA	C	18	48.4	36.8	0.09	2.7	9.75	0.7	0.019	2	54.4	29.9
40	10	FA	C	13	49.4	35.4	0.2	4.7	9.75	0.4	0.045	1¼	54.0	26.1
41	2	FA	C	19	55.4	38.6	0.2	4.5	13.0	0.98	0.019	1	55.4	33.5
42	10	FA	C	10	58.6	38.8	0.06	3.5	18.5	0.7	0.032	1½	54.0	24.3
43	5	FOA	C	15	47.4	39.6	0.09	2.9	15.3	1.6	0.025	1½	54.2	29.5
44	2	FA	C	11	55.6	38.0	0.08	2.3	15.0	1.0	0.019	1¼	54.2	32.6
45	4	FOA	C	9	45.4	37.0	0.2	7.6	7.2	0.5	0.038	2	54.0	28.3
46	5	OA	C	13	59.8	38.0	0.2	3.8	13.0	1.0	0.019	1½	55.4	32.0
47	5	FA	C	11	57.6	36.6	0.06	3.0	18.0	1.5	0.025	1	54.2	27.4
Region 6—North Central														
48	17	WC	GS	10	53.0	39.2	0.09	1.3	19.5	1.3	0.045	2	56.0	24.6
49	1	FOA	C	13	50.8	42.4	0.07	1.0	64.6	3.3	0.032	1½	55.4	32.5
50	5	FA	C	9	58.8	36.4	0.1	1.7	12.6	0.9	0.025	3/4	54.8	27.8
51	4	FA	C	11	45.4	39.8	0.05	0.7	11.8	1.3	0.025	1	55.2	36.0
52	2	FA	C	9	59.4	41.2	0.07	0.4	33.0	2.6	0.019	1½	55.2	31.5
53	5	FA	C	22	54.8	40.4	0.4	5.6	7.2	0.5	0.025	1	54.8	30.9
54	3	OA	C	9	56.2	40.6	0.15	2.3	26.0	1.7	0.006	1½	55.4	29.3
55	2	OA	C	10	40.8	34.0	0.15	2.9	16.3	0.1	0.009	1½	54.8	34.6
56	2	OA	S	14	48.6	39.2	0.08	1.3	33.0	3.25	0.006	1	55.2	33.0
Region 7—East Central														
57	3	FA	C	16	53.4	40.0	0.2	0.9	29.3	1.6	0.025	1½	55.2	38.3
58	3	OA	GS	13	46.0	38.0	0.16	0.7	20.9	2.1	0.013	1	55.4	42.0
59	4	OA	GS	11	45.0	38.4	0.08	2.1	32.5	1.9	0.006	1½	55.0	28.7
60	6	FOA	GS	16	44.4	36.8	0.1	0.8	55.4	3.9	0.025	1½	54.6	39.1
61	4	FA	GS	16	46.4	37.2	0.07	1.3	65.3	4.6	0.013	1½	54.4	38.0
62	1	FA	GS	10	48.6	29.2	0.05	0.6	71.0	7.2	0.006	3/4	55.0	36.7
63	4	OA	GS	12	53.0	36.8	0.09	2.6	13.0	1.3	0.006	1½	56.4	30.2
64	3	FOW	GS	12	45.2	40.4	0.05	0.9	65.0	5.2	0.006	1	54.0	33.3
65	1	FOW	GS	12	45.6	36.8	0.04	0.8	96.0	6.5	0.006	1	54.6	30.0
66	3	FOA	GS	12	50.8	42.0	0.06	0.6	84.9	7.2	0.009	1	54.8	32.5
67	5	OA	GS	13	48.4	23.8	0.2	1.3	16.0	2.1	0.006	1¼	55.2	29.7
68	0	FOW	GS	13	58.2	27.2	0.05	0.5	65.0	6.5	0.006	3/4	54.2	36.2
Region 8—Atlantic Northeast Coast														
69	14	WC	C	12	51.8	32.8	0.3	6.6	6.3	0.3	0.051	2	54.8	22.6
70	1	WC	C	12	52.0	40.6	0.02	0.4	69.3	5.2	0.013	1/2	56.6	33.4
71	14	FA	C	16	51.0	34.2	0.3	1.9	9.2	0.5	0.051	2	53.8	30.2

*OA = self-cooled, FA = forced -air-cooled, FOA = forced-oil-air cooled, FOW = forced-oil-water-cooled, and WC = water-cooled.

†C = conservator, GS = gas-sealed, GOS = gas-oil-sealed, and S = sealed.

‡PPM = parts of water per million of oil, by weight.

Power Factor (ASTM D-924)

"Power factor is the ratio of the power dissipated in the oil in watts, to the product of the effective voltage and current in volt-amperes, when tested with a sinusoidal field under prescribed conditions. It is usually measured at 25°C and reported as per cent power factor measured at 60 cycles corrected to 20°C. This test is an indication of the presence of contaminants or deterioration products such as water, oxidation products, metal soaps, charged colloids, etc."

Resistivity (ASTM D-257)

"The volume resistivity in ohms centimeter of an oil is the ratio of the d-c potential gradient in volts per centimeter paralleling the current flow in amperes per square centimeter at a given instant of time and under prescribed conditions. A low resistivity normally indicates the presence of conductive contaminants but has not been widely applied to used oils."

Water Content

A Karl Fischer electrometric technique previously described[3] was utilized to determine the water contents in this study. This method is presently under consideration as an industry standard by the ASTM for determination of water content of insulating liquids. Water is detected by an electrometric circuit that allows a determination to within ±2 parts of water per million parts of oil by weight.

Acidity or Neutralization Number (ASTM D-664 and D-974)

"Neutralization number for used oils is in general a measure of the acidic constituents of an oil and is the milligrams of KOH required to neutralize one gram of oil. It is a measure of the oil deterioration and some types of contamination."

Color (ASTM D-155)

"The color of an insulating oil is determined by means of transmitted light and is expressed by a numerical value based on comparison with a series of color standards. A rapidly increasing or high color number is an indication of oil deterioration or contamination or both."

Viscosity (ASTM D-88 and D-445)

"The viscosity of an oil is its resistance to uniformly continuous flow, without turbulence, inertia, or other forces. The viscosity of insulating oil is usually measured by the time of flow of a given quantity of oil under controlled conditions. Viscosity is not significantly affected by oil contamination or deterioration but may be useful for identifying certain types of used insulating oils."

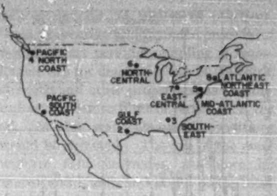


Fig. 1. Sampling regions for field oil survey

Interfacial Tension (ASTM D-971)

"The interfacial tension between an electrical insulating oil and water is a measure of the molecular attractive force between their unlike molecules at the interface. It is expressed in dynes per centimeter. This test provides a means of detecting soluble polar contaminants and products of deterioration."

Results

The results of all chemical, physical, and electrical tests for the 71 transformers are compiled in Table I, arranged by geographic region. The data of primary importance are shown in Figs. 2 through 9, where each symbol plotted represents the average value; the lines extend to show the range between minimum and maximum values. A survey of the data shows no noticeable effect on the properties because of geographic location or type of transformer cooling. Some properties are affected by the method of oil preservation and age of the oil.

Fig. 2 shows that dielectric strength, the most important property, when measured by the VDE cell, is not affected by the age of transformer oil, nor can any trend be seen owing to the type of preservation used. Similarly, no trends can be detected when the ASTM test cup is employed; see Fig. 3.

The water content of the transformer oil is not affected by oil age or method of preservation; see Fig. 4. Of particular importance is the fact that the water content did not exceed 22 parts per million on any unit.

The power factor data, Fig. 5, show a slight increase caused by age and no consistent difference is evident for the different types of preservation. Resistivity of the oil, however, shows a corresponding decrease with oil age, Fig. 6, and, again, all methods of preservation appear to give similar results.

Interfacial tension, as shown in Fig. 7, decreases with age, more in the early years than later, but no effect of type of preservation is noticeable. Similarly, color increases slightly in the first few years, leveling off at a value of approximately two; see Fig. 8. No influence of preservation is seen.

The acidity appears to increase with increased oil life, as shown in Fig. 9. The conservator units appear to have very slightly higher values of acidity, although all values are well below that considered to be acceptable on new oil.

It is interesting that the oil in only six transformers had been filtered and none had been reclaimed during the operating experience of the 71 units. No time pattern or reason was evident in the filtration (units were filtered after an oil life ranging from 4 to 13 years). Table II summarizes the maximum (or minimum) property values obtained on all units and on 95 per cent of all units.

Table II. Summary of Oil Property Value

Property	Minimum Value for All of Units	Minimum Value for 95 Per Cent of Units
VDE dielectric stress, kv rms	36.8	44.4
ASTM dielectric stress, kv rms	23.8	30.8
Interfacial tension, dynes per centimeter	22.0	23.6
Resistivity, 25 degrees centigrade, 10^{11} ohms-centimeter	6.3	9.2
	Maximum Value for All of Units	Maximum Value for 95 Per Cent of Units
Power factor, 25 degrees centigrade, per cent	0.40	0.2
Water, parts per million	22	18
Acidity, milligrams KOH per gram	0.051	0.045
Color (ASTM)	$2^1/_2$	2

Significance of Results

The tests described are those which are generally used to evaluate the serviceability of insulating oils in use in a wide variety of applications. Even when thinking is restricted to transformer applications, the demands placed upon the liquid vary considerably with specific applications and individual power company criteria. In view of this, it is difficult to set arbitrarily maximum or minimum values for the various oil characteristics normally evaluated to determine the continued serviceability of oil in use. It is possible, however, to recognize certain practical limiting values and, in at least one instance, it is possible to cite an acceptable range of values.

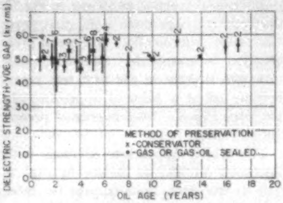


Fig. 2. Effect of transformer oil age on dielectric strength, as determined by the more uniform field-type electrodes. Points are averages, lines indicate spread, and the number of units is shown. (This applies also for Figs. 3 through 9)

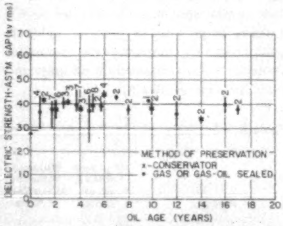


Fig. 3. Effect of transformer oil age on dielectric strength, as determined by ASTM electrodes

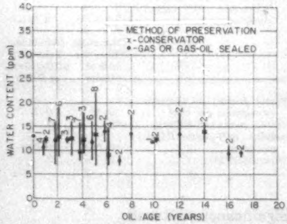


Fig. 4. Effect of transformer oil age on water content of oil

One of the more important criteria used by the power industry to maintain oil in acceptable condition is the ASTM dielectric strength test. Although there is no unanimous agreement on the limiting serviceability values determined by this test procedure, a wide segment of the industry considers a value of 22 kv or higher as indicative that the oil does not need reconditioning or treatment in so far as breakdown strength is concerned. As indicated in this survey, every one of the units exhibited an ASTM dielectric strength value well above 22 kv. In fact, for this oil characteristic alone, all but one of the oils tested would meet a minimum dielectric strength value of 26 kv, which meets the new oil specification of most of the industry. It is interesting to note that 95 per cent of the values reported here have electrical strengths greater than 30 kv.

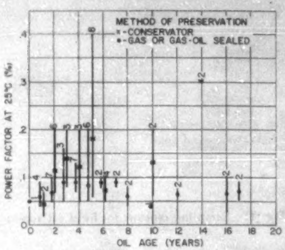


Fig. 5. Effect of transformer oil age on 60-cycle power factor

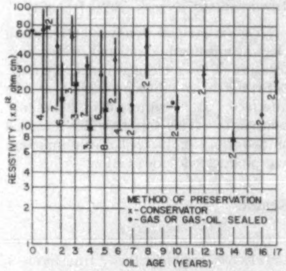


Fig. 6. Effect of transformer oil age on resistivity

In discussing other pertinent electrical, physical, and chemical characteristics which are determined on oils in service, it must be recognized that no single oil characteristic at the present state of knowledge can be relied upon as the sole criterion of oil serviceability. Because of the variety of application conditions, transformer types, the growing trend toward preventive maintenance, and the interdependence of some of these oil tests, it is almost impossible to select a limiting value for any one oil characteristic which would be applicable in all cases. Of greater significance is the evaluation of the general picture established by all pertinent oil characteristics. An estimate of continued serviceability can be intelligently made in any specific case only if the information obtained is focused in terms of the actual unit being considered—its design, rating, age, history, loading, preservation means, and other factors known to the operator.

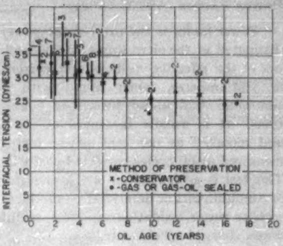


Fig. 7. Effect of transformer oil age on interfacial tension

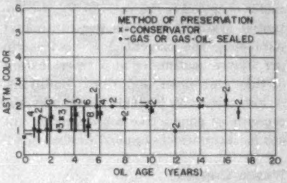


Fig. 8. Effect of transformer oil age on ASTM color

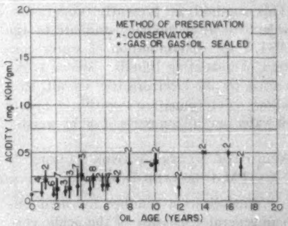


Fig. 9. Effect of transformer oil age on acidity

For the purpose of the present investigation, which deals specifically with larger power transformers embodying three methods of oil preservation, a critical study of the results in Table I indicates beyond doubt that the over-all oil quality of the 71 units surveyed is exceptionally good even by some of the more stringent criteria employed by many utilities. In light of the discussion given earlier, it is felt that these oils can undoubtedly be classified as being in satisfactory condition for continued use. In addition, the data obtained do not indicate that the oils have been seriously affected by the general lack of oil reconditioning nor should any limitation be placed upon their continued use.

Conclusions

The reported survey of oil quality in 71 power transformers reveals that the dielectric strength and all other oil properties are unaffected or only slightly affected by use for periods up to 17 years. Location within the country and type of cooling appear to have no effect on the oil properties, but some properties remain slightly better when preservation is accomplished by gas-sealed or gas-oil-sealed methods.

The effectiveness of the modern oil preservation systems in maintaining the quality of the oil in the transformers studied is emphasized by the apparent lack of periodic oil treatment practices. Good preventive maintenance procedures would dictate periodic sampling of transformer oils for determination of dielectric strength and other properties, as an indication of when to filter or recondition the oil. In this way, abnormal conditions would be detected. The use of the more uniform field electrodes, as shown by Weber and Kaufman,[2] would be the more desirable method for the dielectric strength determination.

The excellent behavior of these oils in use for periods up to 17 years indicates that very little or no extra effort or expense would be required to maintain transformer oils in high-quality condition over extended periods of time, and that degradation of oil in service is not a limiting factor in utilizing the cost-saving advantages of insulation reduction of two full levels.

References

1. Drop Power Transformer BIL 2 Steps, J. R. Meador. *Electrical World*, New York, N. Y., vol. 145, no. 8, Feb. 20, 1956, pp. 103–05.

2. Evaluation of Dielectric Strength Gages for Transformer Oil, K. H. Weber, R. B. Kaufman. *AIEE Transactions*, vol. 75, pt. III, 1956 (*Paper 56-745*).

3. Gas and Moisture Equilibriums in Transformer Oil, R. B. Kaufman, E. J. Shimanski, K. W. MacFadyen. *Ibid.*, vol. 74, pt. I, July 1955, pp. 312–18.

◆

Discussion

L. Wetherill (General Electric Company, Pittsfield, Mass.): The outstanding conclusion which the authors have shown is that oil in power transformers with modern oil preservation features such as automatic gas seal or conservator construction retains very good quality for a great many years. It is reassuring to see a sufficient statistical study to demonstrate that oil is being maintained in quality so well in power transformers. This furnishes a sound basis for concluding that transformer oil can be and generally is maintained in suitable condition for insulation levels reduced two steps. Other factors must, of course, be considered in each specific case before the suitability for insulation level reduced two steps can be determined.

While the authors survey a number of different oil characteristics, perhaps the dielectric strength obtained with the VDE-type electrodes is of outstanding significance in its bearing on the paper by Weber and Kaufman (see reference 2 of the paper) dealing with the development of this same gap.

One of the two primary functions of the oil in a power transformer is to withstand dielectric stress and most of the tests which are used in evaluating oil are aimed directly or indirectly at measuring the ability of the oil to withstand dielectric stress and to continue withstanding dielectric stress. It has always seemed a little surprising that in evaluating the dielectric strength of oil, measurement of dielectric strength has played such a small part and measurement of so many other characteristics has been found necessary. An extreme view which has been expressed sometimes is that the dielectric strength of the oil is the one characteristic which does not have to be measured because the other tests customarily made on oil are more sensitive indicators of suitability of oil for use in transformers than the dielectric strength test. The data presented by the author throw no light on this situation, but the paper by Weber and Kaufman indicates that if the test results obtained on the VDE-type gap are considered the situation is considerably changed, since the dielectric strength measurements are indicative of dielectric strength as demonstrated in a transformer. This should remove some of the present dependence on other tests to indicate quality of insulating oil.

The authors are to be congratulated on having added to our understanding of dielectric strength characteristics of oil in service and how it is best measured. It is to be hoped that others will conduct investigations in this same area so that the industry can benefit through proper utilization of a rounded contour test gap.

W. J. Degnan and **E. J. Shimanski**: We are grateful to Mr. Wetherill for his comments. The finding that modern oil preservation methods allow successful maintenance of transformer oil quality has led to further study.

About 5 years ago 12 transformers were constructed which incorporated insulation reduced two levels; these units have been in successful operation since that time. An oil survey of these transformers has been conducted since preparation of this paper. Tests, identical to those covered in the paper, were made on samples from these units and in all cases the values obtained were in complete agreement with the data presented in the paper.

This is added evidence that degradation of oil in service is not a limiting factor in reducing insulation two insulation levels.

Some of the Engineering Aspects of the Machine Translation of Language

ROBERT E. WALL, JR.
ASSOCIATE MEMBER AIEE

THE MECHANICAL translation of language by digital data-processing equipment was first suggested in private correspondence between A. D. Booth and W. Weaver in 1946.[1] Since that time, others have been attracted to the field of mechanical translation and their work has resulted in the publication of a magazine, a book, and numerous papers.[2,3,4]

At the time that mechanical translation was conceived by Weaver and Booth, the necessary computer components were not available to realize a mechanical translator. In this paper it will be shown that the recent advances in computer technology have made mechanical translation a definite possibility provided certain limitations are made in the type of material which will be fed into the machine, and also if certain allowances are made in the required quality of the output material.

The intent of this paper is to acquaint engineers with the general "state of the art" of the machine translation problem and to indicate some of the more promising methods of attack on the individual problems. The individual problems are not considered in detail. A careful analysis of either the coding problem or the memory requirements would require more space than is used for this entire paper. Detailed discussions will be handled in subsequent papers.

General Considerations

The people of every highly civilized speech community have a rich literary heritage. It would not seem reasonable to expect the first translating machines to translate this literature from one language to another and to preserve completely the original style and the more delicate shades of thought and meaning. There is, however, a very heavy demand for translations of scientific material at present. Since scientists tend to use a simple language structure in their writing, and since readers of scientific material are interested mainly in the intelligibility of the material, it is possible to make many useful restrictions on the amount of processing which a machine translator would have to perform if the input is restricted to scientific writing. For example, it has been shown[5] that Russian scientific writers tend to follow English word order form. It is likely that a machine translator can produce satisfactory Russian-English translations without rearranging word order in the "target" language. The "target" language is defined as the language into which the translation is made, while "source" language refers to the language from which the translation is made. The fact that only intelligibility in the output material is required allows a considerable number of restrictions in the number of target language words stored in the dictionary. For example, a Russian-English dictionary[6] lists with the word "окаянство" the English equivalents "impiety, impiousness, ungodliness, godlessness." Nothing essential to understanding would be lost if only one of the four equivalents were stored. It might also be advisable to simplify the spelling of English words to decrease storage space, for example by storing "thru" instead of "through."

It seems unlikely that the quality of the translations from a machine would be equal to that supplied by a well-trained human translator. The machine would have to justify its existence by superiority over the human in speed and, if possible, by a lower per-word cost of translation. A language translation machine would undoubtedly represent a very large investment since it would require a large temporary storage space to allow the necessary processing of material, and it would have to have a permanent memory with a very large capacity and a relatively low access time. It seems likely that such a machine will cost as much as a modern large-sized digital computer which has a per-hour cost of about $250. Human translators will produce rough, understandable translations for a per-word cost of about one cent. This means that a $250-per-hour machine would have to produce about 25,000 words per hour or, in round numbers, about ten words per second, in order to have the same per-word cost. This figure is, of course, somewhat arbitrary, but it is certain that 1 word per second is too slow, and 100 words per second is probably faster than necessary.

Translation Process

Four definite steps are required in the translation process: encoding, memory search, logical operations, and decoding. Consider these four steps in order.

ENCODING

Two basic problems arise when an attempt is made to encode text material into the machine language, or "code." The first problem is that of the optimum selection of the machine code, and the second is that of transcribing printed text material into machine code. Printed material could be converted into machine code in many ways. Operators might transcribe the text material onto punched or magnetic tape; an electronic tape reader could then read the tape and supply the necessary electrical pulse patterns to the machine. As another possibility, the duties of the human transcriber and the mechanical tape reader might both be performed by an all-electronic device. A machine has been recently announced which will read and encode English at the rate of 3,600 words per minute.[7] It should be possible to modify this machine so that it will read and encode the Russian Cyrillic alphabet. It is beyond the scope of this paper to discuss the configuration of an electronic encoder in detail; however, such a device is very important for the realization of economical machine translation since human transcribers who know only English would copy foreign language material slowly and there would be a high probability of error.

The machine code should be selected in such a way that the code groups were of uniform length, and so that storage of material would require the least number of bits. It appears that it is necessary to provide both capital and lower-case letter codes in order to process the material properly. The total requirement for Russian will then be about 80 distinct code groups: 9 for the number codes, about 10 for punctuation marks, and 61 for letter codes (5 of the 33 Russian letters never start a word and will not

---

Paper 56-693, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 27, 1956; made available for printing May 4, 1956.

ROBERT E. WALL, JR., is with the University of Washington, Seattle, Wash.

Gratitude is expressed to Profs. Laurel J. Lewis, W. Ryland Hill, and Erwin Reifler for many helpful criticisms and discussions.

Table I. Proposed Code

Symbol	Lower Case Code	Capital Code
а	00001	1111100001
б	00010	1111100010
в	00011	1111100011
г	00100	1111100100
д	00101	1111100101
е	00110	1111100110
ё	1111011110	
ж	00111	1111100111
з	01000	1111101000
и	01001	1111101001
й	1111111010	
к	01101	1111101010
л	01011	1111101011
м	01100	1111101100
н	01101	1111101101
о	01110	1111101110
П	01111	1111101111
р	10000	1111110000
с	10001	1111110001
т	10010	1111110010
у	10011	1111110011
ф	10100	1111110100
х	10101	1111110101
ц	10110	1111110110
ч	10111	1111110111
ш	11000	1111111000
щ	11001	1111111001
ъ	1111000010	
ы	11010	
ь	1111000011	
э	11011	1111111011
ю	11100	1111111100
я	11101	1111111101
1	1111000100	
2	1111000101	
3	1111000110	
4	1111000111	
5	1111001000	
6	1111001001	
7	1111001010	
8	1111001011	
9	1111001100	
Space	00000	
Comma	1111001110	
Semicolon	1111010000	
Period	1111010001	
etc.		

need to be capitalized). A demand for 80 distinct code groups would seem to require a 7-bit code, but it will be shown that a basic 5-bit code may be made to yield 80 distinct code groups and, if properly selected, the 5-bit code will result in a considerable saving in storage. An example of such a basic 5-bit code is displayed in Table I. To test this code a random sample of a half-page of Russian text was analyzed; the occurrences of the various symbols are listed in Table II. The sample had a total of 786 symbols. Coding this material with a 7-bit code would require 786×7 totals 5,502 bits. If the code shown in Table I were used, there would be 745 occurrences of 5-bit codes and 41 occurrences of 10-bit codes; see Table II. The total number of bits required would be $41\times10+745\times5=4{,}135$ bits, or an average of $4135/786=5.23$ bits per symbol. Thus a considerable saving in storage space results from using this modified 5-bit code over that required with a 7-bit code.

Many other codes could be devised. The illustrated code does have a uniform code-group length. It was shown that a random sample of text could be coded with an average code length per symbol of only 5.23 bits; this is only 75 per cent of the bits required with a 7-bit code.

MEMORY SEARCH

The permanent memory system of the mechanical translator must store essentially all source language words except for those words (internationals) shared by the source and target languages. The permanent memory must also store essentially all source language word sequences (idioms) which do not give understandable word-for-word translations into the target language. Each source language entry must have the necessary target language equivalents stored with that entry in the memory. During the memory search, incoming text words may be compared with the source language entries in the memory, first in groups of words, with the idiomatic memory, then by single words, with the individual-word memory.

It has been shown[8] that words of rare occurrence must be translated, since very often these words determine the meaning of the whole sentence. Therefore, it would seem necessary to store the complete source language-target language dictionary, with the exception of internationals. The largest German-English dictionary[9] contains approximately 400,000 entries. When the selection of English equivalents is made it should not be necessary to store more than an average of two equivalents for each source langage words. If internationals are not stored, the total storage requirement is probably about 1,000,000 words. A reasonable estimate of the average length of a word would be 7 letters, or a little more than 35 bits using the code of Table I. The individual-word memory would thus contain about 40 megabits of information. The storage requirement for the idiomatic memory is considerably less, definitely not over 60,000 words in total.

The Russian language does not have as many words as either German or English but the storage requirements are probably about the same because in Russian 59 word affixes[10] are used to denote the grammatical usage of a word. The Russian noun has six cases and may be either singular or plural, and an adjective must agree with the noun that it modifies both in gender (there are three genders in Russian) and in case.

Table II. Occurrence of Russian Letters, 1/2 Page of Text

Symbol	No. of Times Occurring
а	35
б	5
в	41
г	7
д	12
е	57
ё	4
ж	2
з	12
и	79
й	6
к	24
л	26
м	18
н	50
о	63
П	19
р	42
с	36
т	37
у	15
ф	10
х	14
ц	11
ч	12
ш	1
щ	2
ъ	0
ы	17
ь	4
э	5
ю	5
я	17
Space	80
Comma	11
Colon	0
Semi colon	0
Period	4
Nos. and Symbols	5
Capitals	5

The amount of material stored in the memory system should be kept as small as possible, and the material should be stored in the memory in such a way that the search may proceed in the most efficient manner. It is possible to decrease the amount of material stored in the memory by not storing internationals. Other schemes for decreasing dictionary size depend on reducing a word to simpler components, such as reducing a compounded word to the words of which it consists, or dissecting a word into a stem and an ending in the case of affixed words. Eliminating the storage of internationals presents no serious logical difficulties per se. If an incoming text word should not correlate with any of the source language words in the dictionary, the word could be printed out in its original form or in its target-language phonetic equivalent. It is necessary to make a provision for printing out uncorrelated words anyway, in order to take care of misspelled words. Routines which involve dissection schemes would give a considerable decrease in memory size, but they all pre-

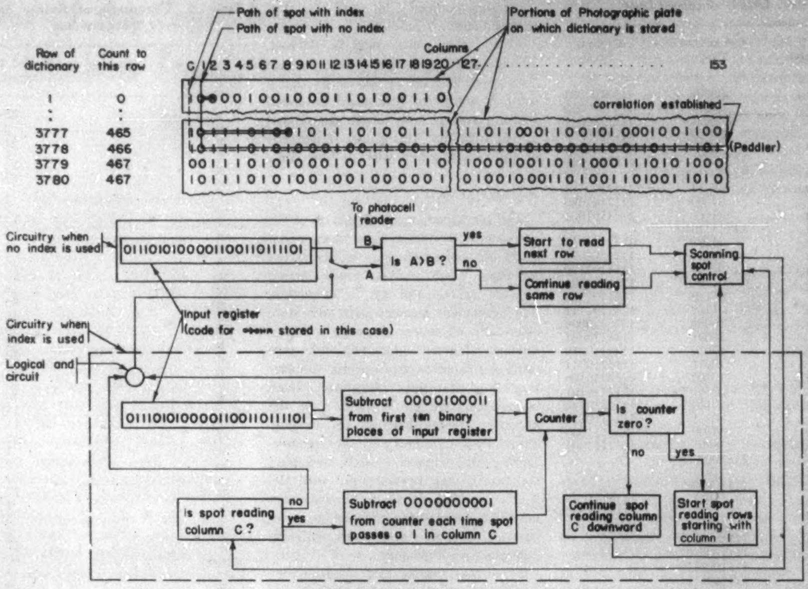

Path of spot with index
Path of spot with no index
Portions of Photographic plate on which dictionary is stored
Columns
Row of dictionary
Count to this row
C 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 127 153
1 0
3777 465
3778 466
3779 467
3780 467
correlation established
(Peddler)
To photocell reader
B
A
Is A>B ?
yes
no
Start to read next row
Continue reading same row
Scanning spot control
Circuitry when no index is used
01110101000011001101110l
Input register
(code for [illegible] stored in this case)
Circuitry when index is used
Logical and circuit
01110101000011001101110l
Subtract 0000100011 from first ten binary places of input register
Counter
Is counter zero?
no
yes
Continue spot reading column C downward
Start spot reading rows starting with column 1
Is spot reading column C ?
no
yes
Subtract 0000000001 from counter each time spot passes a 1 in column C

Fig. 1. Block diagram of optical memory

sent logical difficulties in their complete form. For instance, if separate stem and ending dictionaries are provided, difficulties result from the many irregularities of language and the necessity of coordinating the two dictionaries. It would seem useful to break up a word into stem and ending within the individual dictionary entry by listing the word stem first then the endings along with the target language equivalents. An incoming text word may then be compared with a stored source language word starting with letters farthest to the left in the words. If correlation is established, the target language equivalents stored with the correlated ending are read out and stored to await further processing.

A scheme for dissecting compounds has been developed by Reifler.[11] If a word is not located in the dictionary, the letter at the end of the word is ignored and the remainder of the word is compared with the dictionary. If this remainder is not located, another letter at the end is ignored. This process is continued until the remainder is located or until all the letters have been used up. Then the machine starts to ignore letters from the front of the word and to compare the remainder with the dictionary as before. When a portion of an incoming word has correlated with a stored sequence, then the ignored portion of the word must also be compared with the dictionary. If letters are ignored from the front of the word, or if ignored portions of a word are to be compared, it often would be necessary to go to a part of the memory which is far removed from the place that was searched in the previous step. This skipping about in a large memory system is a very serious problem. Any word which could not be correlated with a word in the dictionary would have to undergo any compound dissection routine which was built into the machine; thus all misspelled words and internationals would have to be tested for compounding before they were printed out. Since about ten per cent of ordinary Russian scientific text is composed of internationals, the time consumed in attempts at compound dissection would probably be prohibitive. If only letters at the end of the word were ignored, the ignored portion has not been compared with the memory, and if the routine were halted when all the letters have been used up, then no appreciable skipping about the dictionary would be required. This much of the compound dissection scheme might be profitably used.

There are many possible variations for the order in which words might be stored in the dictionary. They might be stored first according to the length of the word, then alphabetically; the words might be stored first according to grammatical type, then alphabetically; or the storage might be completely alphabetical. Each of these systems has certain advantages, but complete alphabetization has one clear-cut advantage over all other methods; each word then has a unique address in the store determined by one criterion alone, the spelling of the word. In this paper, only complete alphabetization of the dictionary will be considered.

The permanent memory has two design limits: a maximum allowable access time, and a maximum allowable size. The maximum size of a large permanent memory is mainly determined by physical limitations; for instance, a memory using a photographic plate to store the information may be limited in size by the area of plate which can be scanned. If a particular memory of the photographic-plate type is capable of storing 50 megabits of information, the memory cost will be approximately the same whether it stores 50 megabits or 15. Access time is actually determined by the efficiency of the search routine. Thus it is sometimes better to store more material in order to realize a more efficient search routine. The first criterion for determining the amount of material which must be stored in the memory system is necessarily the linguistic requirements of the problem. Once the linguistic requirements are satisfied, the all-important criterion for determining size of storage, sequence of storage, and the method of coding the material is the effect of each on the access time into the memory. The physical configuration of the memory largely establishes the optimum storage sequence and the most efficient indexing procedure. A likely device for the large memory of the translation machine is an optical device[12] in which the information would be stored on a photographic plate and the plate scanned by a beam of light formed by a cathode-ray spot and focused by a lens. To store the required 40 megabits of information in a memory of this type there might be about 6,500 rows of about 6,500 bits each. This would give a total of 42.5 megabits of storage. The discussion of storage sequence and indexing will be confined to schemes applicable to this type of storage. A hypothetical portion of a photographic plate on which a part of a dictionary is stored is illustrated in Fig. 1. Actually the 1's would most likely be just spots on the plate and the 0's just a lack of a spot, but the representation shown is easier to comprehend.

It can be shown that it is important for the largest part of an incoming text word to be correlated first. For instance, if the source language is English, incoming text words should be compared with "manage" before "man," and with "indicated" before "in." For this reason the source language entries shown in Fig. 1 are stored in inverse numerical sequence in a row, i.e., the numerically largest code group in a particular row is stored to the extreme left of the row with its target language entries, then the numerically next largest source language code group, and so on. The entries in column *C* of Fig. 1 are indexing entries which will be explained later. The word codes are entered in the rows starting with column 1, first the source language word code and then its target language equivalents. In Fig. 1, row 3780 stores the code for "офталмоскоп" as its first source language word, and row 3779 stores "офицновый" as its first source language word, etc. The word which will serve as an example is "офеня" and it is assumed to be stored in columns 127–153 of row 3778.

If the incoming text word were "офеня," and no indexing scheme were used, the spot would move in a path traced by the solid line in Fig. 1. The circuitry enclosed by the solid line would be used in this case. In each row, in order, the spot would read the code until a binary place was reached which was different from the corresponding binary place in the text-word code. If at that binary place the text code had a "1" and the dictionary code had a "0," the spot would start to read the next row. If the reverse were true, the word must be located in that row if it is stored in the dictionary. The spot would continue reading that row until correlation was established or until a stored source language code sequence had a smaller code number than the text word, which would establish the fact that the text code was not stored in the dictionary. A dissection routine would then be started if such schemes were included in the stored logic. With this search procedure, the spot must read an average of perhaps three or four binary places at the beginning of each row before the machine could determine whether or not it should go to the next row or should continue to read the same row. An indexing scheme may be used to reduce the average number of binary places that are read in a row before the machine would decide whether or not it must go to the next row. This is particularly easy in this example since complete alphabetization of the dictionary is assumed.

The circuitry necessary for one possible indexing scheme is represented in block form in Fig. 1, enclosed in a dashed line. This scheme depends upon the fact that the first two letters of a word will determine closely the row in which the word must be stored. A "1" is placed in column *C* in a particular row each time the first two letters of the first word stored in that row are different from the first two letters of the first word stored in the previous row. The idea is to treat the first two 5-bit code groups of an incoming text word code as a binary number; the rows which start with "оф" (code 0111010100 which is binary for decimal 468) then form row group 468. For example, suppose that the incoming text word is "офеня" (peddler). The first two letters, "оф" will have the code 0111010100.

This code sequence is transferred to the reversible binary counter, leaving the contents of the register unchanged. The spot then moves down column *C*. subtracting a binary "1" (0000000001) from the counter each time the spot passes a "1" in column *C*. When the content of the counter is reduced to zero, the machine starts reading the rows, as indicated by the dashed line in Fig. 1, exactly in the same way as when no indexing scheme was used. Two complications arise in the scheme: first, the first row of the dictionary starts with 0000100100 (it is assumed that аГенЦ (lamb) is stored in this position), which is considerably larger than 000000001 and, second, some letter sequences do not exist as the first two letters of a word, for instance, there are no words beginning оу . . . . The first complication is resolved by the use of a subtractor between the input register and the counter which automatically subtracts a binary 0000100011 from the 10-bit sequence as it is transferred to the counter. If a negative difference is obtained when this subtraction is performed, the word must be located in the first row. If the difference is positive, the spot moves down column *C* subtracting 0000000001 from the contents of the counter each time it passes a binary "1." The second complication is resolved by placing dummy "1's" in column *C* in rows which have the same first two letters as the preceding row. For instance, alphabetically (and numerically for the corresponding codes) the sequences are "от, оу, оф," but no "оу . . ." words exist. Therefore, to keep the count straight a dummy "1" is placed in column 1, row 3777. The search for the word "офеня" would then follow the path shown by the dashed line. This path would require reading only one binary place per row until the spot had approached within an average of six or seven rows of the desired row.

About 1,200 words[13] constitute approximately 40 per cent of text material in most languages so it would seem profitable to include these words of short length and very high probability of occurrence in a separate section of the memory. Then incoming text words could be compared with this small dictionary before comparison with the large memory. This auxiliary small dictionary would be used if the other machine operations take substantially less time than the dictionary search, since its use would undoubtedly speed up the average location time of words. If the dictionary search was fast enough so that the other machine operations were the limiting factors, then the logical complications of two single word dictionaries would not justify splitting the dictionary.

Idioms could be stored in a separate dictionary and incoming words compared with this dictionary before comparisons with the single-entry dictionary. These idioms might also be stored in the same dictionary as the single-word entries. Storing both the idioms and single-word entries in the same dictionary would probably give a lower average location time, but the logical problems would be more complex. The operation of a separate idiomatic dictionary will not be discussed in detail since source language text material could be correlated with this dictionary in the same manner as with the individual-word dictionary although the routine would be somewhat more complicated. This idiomatic dictionary should be small compared with the individual word dictionary.

To store 40 megabits of information on a photographic place in the manner shown in Fig. 1, a square of information about 6,500 bits on a side is required, which yields a total storage of about 4.25×10^7 bits. With 6,500 rows in the large dictionary, the average number of rows passed by the spot before it gets to the proper section is about 3,250. Since the rows also have about 6,500 bits stored in them, the average number of bits read in the proper row would be about 3,250, or a total of about 6,500 bit positions to read, on the average, before a word is located in the large dictionary. If a repetition rate of 100 kc is assumed (this will allow relatively simple transistor circuitry to be used) the average location time for an incoming text word would be about 0.065 second. Thus it seems possible to construct an optical store which would have the required capacity and allow an access time of less than 1/10 second, which was previously estimated as an economical speed of operation.

LOGICAL OPERATIONS

Word-for-word translation has proved surprisingly good, but not satisfactory for commercial translation. Some consideration of the context of words seems necessary to obtain a satisfactory output. Yngve[14] suggests consideration of perhaps three or four words on each side of a particular word when attempting to determine a unique meaning from a set of multiple meanings of the word. He does not suggest the consideration of words not located in the same sentence as the word being processed. To aid in such logical operations, linguistic "tags,"[13] or code sequences which denote certain grammatical information about the word, may be stored in the memory along with the target language equivalents of the word. Every source language word having more than one target language equivalents would require the tag. At present, it is not possible to estimate how much processing would be required, and how much information and of what type should be stored in these tags since linguistic research has not progressed very far in this direction. Some work has been accomplished[13] on grammatical incident meaning but nothing has been done on the nongrammatical incident meaning of words. The logical operations are probably the most important area for contemporary research. To illustrate how the linguistic tags might be used to eliminate impossible target-language equivalents, the following example given by Reifler (page 152 of reference 13) is reproduced. Consider the German expression "wegen dieser Schueler" (because of these pupils). "Dieser" has in isolation the possible grammatical meanings of singular masculine nominative, singular feminine genitive or dative, or plural genitive. The word "Schueler" has in isolation the grammatical meanings of singular masculine nominative, dative, or accusative, or plural nominative, genitive or accusative. Consider the following tag assignments.

1. Singular masculine nominative.
2. Singular feminine genitive.
3. Singular feminine dative.
4. Plural genitive.
5. Singular masculine dative.
6. Singular masculine accusative.
7. Plural nominative.
8. Plural accusative.
9. Singular genitive.

The tags stored with "wegen" will be 4, 9; with "dieser," 1, 2, 3, 4; and with "Schueler," 1, 4, 5, 6, 7, 8. The cooccurrence of "wegen," "dieser," and "Schueler" narrows down the two alternatives of "wegen," the four alternatives of "dieser," and the six alternatives of "Schueler" to only one possibility, that plural genitive since only tag 4 is common to all three.

DECODING

In the decoding operation, material in machine code would be translated in some way into the target-language alphabet. A device has been recently reported which accepts code and translates it into printed form on a cathode-ray screen at the rate of several thousand letters per second.[5] The material can then be photographed, and erased, and new material can be introduced. This printing speed is an order of magnitude greater than the minimum that is required for a translator.

Conclusions

There appear to be no insurmountable engineering problems in the realization of a mechanical translation machine. Research has progressed sufficiently to allow the formulation of a detailed program which should result in a commercially practical machine in a reasonable time. Much work remains to be done, most of which is linguistic in nature, but a substantial number of good engineering problems remain to be solved.

(*Discussion follows.*)

References

1. Machine Translation of Language (book), **A. Donald Booth, William N. Locke**, editors. "Historical Introduction," **Booth and Locke**. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 2 ff.

2. *Mechanical Translation* (periodical), Massachusetts Institute of Technology, Cambridge, Mass.

3. **Booth and Locke**, *op. cit.*

4. Computing Machines for Language Translation, **Thomas M. Stout**. *The Trend in Engineering*, University of Washington Press, Seattle, Wash., vol. 1, no. 3, Dec. 1954, pp. 41–46.

5. The Mechanical Translation of Russian: A Preliminary Report, **Kenneth E. Harper**. University of California, Los Angeles, Calif.

6. Russian-English Dictionary, **Louis Segal**. G. E. Stechert and Company, New York, N. Y., 1942.

7. Photoelectric Reader Feeds Business Machines, **David H. Shepard, Clyde C. Heasly, Jr.** *Electronics*, New York, N. Y., vol. 28, no 5, May 1955, p. 134.

8. **Booth and Locke**, *op. cit.* "Some Problems of the Word," **William E. Bull, Charles Africa, Daniel Teichroew.**

9. Enzyklopädisches Englisch-Deutsches und Deutsch-Englisches Wörterbuch, **Muret-Saunders.** Langenscheidtsche Verlagsbuchhandlung. G. m. b. H. Berlin-Schöneberg, Germany, 1910.

10. **Booth and Locke**, *op. cit.* "The Design of an Automatic Russian-English Dictionary," **Anthony G. Oettinger.**

11. Mechanical Determination of the Constituents of German Compounds, **Erwin Reifler.** *Mechanical Translation*, Massachusetts Institute of Technology, Cambridge, Mass., vol. 2, no. 1, 1955, p. 3.

12. Photographic Techniques for Information Storage, **Gilbert W. King, G. W. Brown, L. N. Ridenour.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, no. 10, Oct. 1953, p. 134.

13. **Booth and Locke**, *op. cit.* "The Mechanical Determination of Meaning," **Erwin Reifler.** pp. 159 ff.

14. **Booth and Locke**, *op. cit.* "Syntax and the Problem of Multiple Meaning," **Victor H. Yngve.**

15. An Electron Tube for High Speed Teleprinting, **W. H. Bliss, J. E. Ruedy.** *RCA Review*, Princeton, N. J., vol. XVI, no. 1, Mar. 1955, p. 5.

Discussion

M. M. Astrahan (International Business Machines Corporation, San Jose, Calif.): The 42.5-megabit storage unit proposed would be impractical with presently available or contemplated components. As pointed out in reference 12 of the paper, a cathode-ray tube with a resolution of at least 6,500 lines would be required. Also the problems of accuracy and stability in addressing would be prohibitive.

I would think that the use of an indexing system could considerably reduce the amount of sequential scanning required so that memories with smaller sized high-speed sequential access sections, and slower access to each of the sections, could be employed.

Robert E. Wall, Jr.: Mr. Astrahan is perfectly correct in stating that an indexing scheme may be used to allow rapid processing of material with a memory system, as he describes. Such memories are available now, while the memories of the type indicated in this paper are not now actually available. Until very recently, a Clevite-Brush TapeDRUM was considered for dictionary storage and was to be used as follows: Incoming text material was to be stored in a temporary memory in blocks of about 1,000 words, then each word was to be indexed according to its alphabetical order. The dictionary would then have been searched in alphabetical order for the entire 1,000 words at one time. It is possible to obtain a processing time of ten words per second with such a scheme. The disadvantage of this method is that considerable temporary storage space is required to store and index the incoming material; considerably more than would be necessary just for the logical processing alone. It can also be shown that any type of word dissection scheme is rendered very difficult with a dictionary stored on a low random-access memory. Any processing in the source language, such as perhaps would be required for processing the negations of French, is also much more difficult to perform.

Magnetic Amplifier Control of Switching Transistors

H. WILLIAM COLLINS
ASSOCIATE MEMBER AIEE

Synopsis: A series of circuits are presented which use a magnetic amplifier to control the average output power of switching transistors by pulse-width modulation. Thus the advantages of switched-transistor operation are incorporated in a linear amplifier. The magnetic elements serve both as a pulse-width modulator and as a high-gain preamplifier, thereby producing an amplifier with high-gain as well as high-output power. Design considerations, performance data, and potential applications of the circuitry are discussed.

THE ADVANTAGES of operating a transistor in a switching mode have been clearly demonstrated.[1] These advantages can be briefly reviewed by reference to Fig. 1 which shows the characteristics of a power transistor operated in the common emitter connection. When operated in the switching mode the transistor remains either at point *A*, corresponding to high collector current and very-low collector voltage, or at point *B*, corresponding to high collector voltage and very-low collector current. The power dissipation in the transistor at each of these points is much less than the load power which is controlled by the switching action. In contrast, when operating a transistor in class *A* manner, the internal dissipation equals or exceeds the load power. For a given internal dissipation, the power-handling capability of a transistor, therefore, can be increased typically by a factor of 10 by operating in the switched mode. A corresponding increase in efficiency also results. Because the transistor is operated only at saturation or cutoff, nonlinearities of the transistor-control characteristics do not affect operation. Likewise, very stable operation is achieved because normal variations in the transistor parameters, from aging or production inconsistencies, do not affect circuit operation. Stable operation over a wider ambient temperature range is possible. These features have been exploited in a number of recently developed devices[2,3,4] which use the transistor as a switch.

The transistor can provide an excellent high-speed, static switch which has both high output power and reasonably high power gain. The advantages of the switched mode of operation are similarly attractive in constructing a linear power amplifier, but a means of proportional control of the output must be provided. The average output power of a switching transistor can be linearly controlled if the ratio of "on" time to "off" time is varied as shown in Fig. 2. This method of control is termed pulse-width modulation. A group of circuits has been given which operates transistors in the switched mode in power amplifiers.[5] This paper presents a different technique which uses an audio-frequency magnetic amplifier as a means of pulse-width modulating of power transistors. It will be noted that the technique applied in all of these circuits uses a magnetic amplifier not only as a means of pulse-width modulation, but also as a high-gain preamplifier stage. The resultant circuitry is thereby capable of both high power gain and linear control of the output.

Magnetic Amplifier Control

A self-saturating magnetic amplifier is characterized by sharp saturation of the magnetic-core material which transfers the supply voltage from the magnetic core to the amplifier load.[6] By varying the time in the cycle that saturation occurs (i. e., the firing angle), the output power is controlled. An ideal magnetic amplifier has a control characteristic as shown in Fig. 3. The average output voltage varies linearly with the control signal. If a magnetic amplifier is operated from a square-wave a-c supply, the firing angle also varies linearly between 0 and 180 degrees thus producing a variable-width pulse as shown in Fig. 2. The resultant pulse is ideal as a means of linearly controlling the average output of a power-transistor switching stage.

Conventional magnetic amplifiers are operated from power-frequency sinusoidal supplies. The firing angle of such an amplifier varies sinusoidally rather than linearly with control, and thus is not well suited to controlling a transistor stage which operates from a smooth d-c supply. The basic performance of a magnetic

---

Paper 56-726, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, June 25–29, 1956. Manuscript submitted March 28, 1956; made available for printing May 24, 1956.

H. WILLIAM COLLINS is with the Westinghouse Electric Corporation, East Pittsburgh, Pa.

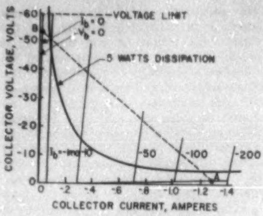


Fig. 1. Common emitter characteristic of power transistor

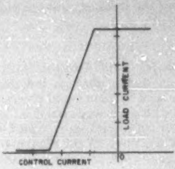


Fig. 3 (left). Magnetic amplifier control characteristic

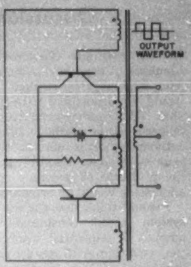


Fig. 4 (right). Saturable core transistor inverter

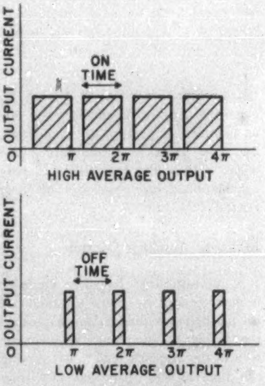


Fig. 2. Pulse-width-modulated output

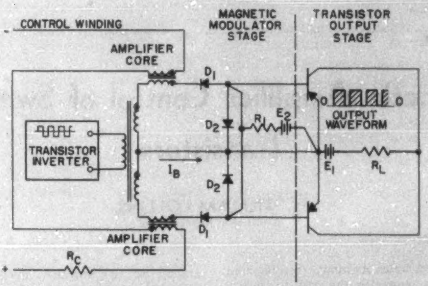


Fig. 5. Basic magnetic-transistor amplifier circuit

amplifier can be greatly increased by operating at high-supply frequencies.[7] Practical high-frequency magnetic amplifiers have been developed[8,9] which have the a-c power supplied by a saturable-core transistor inverter (see Fig. 4). These amplifiers are not only smaller than power-frequency devices, but also have power gains and bandwidths that are normally unattainable in the power-frequency range. Of particular importance in this discussion is the fact that the transistor inverter is a square-wave a-c supply. Therefore, a high-frequency magnetic amplifier has an output pulse whose width varies linearly with control. The magnetic amplifier thus becomes a practical means of linearly controlling switched transistors.

A basic circuit for magnetic-amplifier control of switching transistors is shown in Fig. 5. The magnetic-amplifier stage is a standard center-tapped type of circuit operating at an audio frequency. The power for the magnetic stage is supplied by a center-tapped winding on the saturable core of a transistor inverter. The magnetic-amplifier load consists of two transistors which are pulse-width modulated and their conduction time is linearly varied from 0 to 100% (per cent) by the amplifier control current. In simplest form, the circuit could operate without rectifiers D_1 and D_2, resistor R_1, and voltage E_2. The transistor base to emitter junction can serve as the self-saturating rectifier D_1 for the magnetic amplifier. On the other hand, the leakage of commercially available power transistors is relatively high, and it appears advisable at the present state of amplifier development to use a separate rectifier to assure high gain, linearity, and stable operation of the magnetic stage. A serious problem arises in driving switching transistors from a magnetic amplifier because the magnetizing current (typically several milliamperes) which flows at minimum output allows the transistors to conduct. The situation is greatly aggravated by the high current gain of power transistors at low values of base drive (see Fig. 1). The transistor stage not only cannot be turned off completely but remains operating in the high dissipation region. At high output powers there is great danger of thermal runaway. This problem can be overcome by using biased rectifiers D_2 which bypass the magnetizing current and maintain a positive bias on the transistor base until the magnetic amplifier fires. Isolated d-c source E_1 can be conveniently supplied by a center-tapped rectifier circuit operating from the saturable-core inverter.

Power Gain

The magnetic amplifier serves a dual function as a pulse-width modulator and as preamplifier. In order to evaluate the over-all circuit performance, it is necessary to determine the gain and response of the magnetic stage. The gain and response of a full-wave magnetic amplifier are related by the figure of merit (M) which, for a given amplifier design, is constant over a wide range of response time.

$$M = \frac{A_p}{T}$$

M = figure of merit
A_p = power grain
T = time for 63% response in seconds

The figure of merit of an audio-frequency magnetic amplifier[8,9] is in the range of 10^6 to 10^7. The response time will range from a minimum value of 1/2 cycle of the supply frequency to a maximum value set by the maximum practical ratio of control turns squared (N_c^2) to control circuit resistance (R_c). For a 10-kc magnetic amplifier the time delay may in practice be varied from 0.05 millisecond to about 50 milliseconds by changing the control-circuit resistance. With a figure of merit of 10^7, these delays yield a range of maximum power gains respectively from about 500 to 500,000. In applications where fast response is not required, very-high power gains can be achieved in the magnetic stage. Where greater speed is required, it can be obtained at the expense of increased control power and lower gain. A given combination of time delay and power gain can be effected in a magnetic amplifier at various values of input impedance by appropriately selecting the number of control turns. Thus, this circuitry has an innate means of efficiently coupling to a wide range of source impedances.

The power gain possible with a switching transistor stage is highest at low values of base drive, and decreases as high values of base drive are required. The full signal-power gain of presently available power transistors, operated in a d-c switching circuit with common emitter connection, is about 100. The overall gain of the circuit of Fig. 5 is approximately the multiple of the magnetic- and the transistor-stage gain. Thus, power gains from about 50,000 to 50 × 10^6 are within the range of the circuitry. The voltage gain can be widely altered by selection of the desired input impedance. The magnetic amplifier is basically a low-impedance device, and input resistances greater than several thousand ohms are normally difficult to obtain. Special techniques have been devised, though, which can provide very high steady-state input impedances.[10] It will be noted that the magnetic amplifier normally provides the major power gain of this circuit, and that the switched transistors function primarily as a power-output stage.

Switching Rate

Usually it is desirable to operate these circuits at a high frequency in order to maximize magnetic-amplifier performance and minimize time delays. The maximum practical switching rate is set by the time required for a switching transistor to pass through the high-dissipation region from saturation to cutoff. As this time becomes a significant portion of a half-cycle of the carrier frequency, the average transistor dissipation rises sharply, and the major advantages of the switched operation are lost.

The switching time of presently available power transistors, when maximum collector current in common emitter connection is handled, is about 15μsec (microseconds). The off transient is somewhat longer than the on transient because of hole storage which occurs during the conduction period. The power dissipation of a transistor during the switching period is as follows:[1]

$$P_T = P_L \frac{fT}{3}$$

P_T = average transient transistor dissipation
P_L = maximum load power
f = pulse repetition rate
T = switching time

When switched to maximum collector current, a transistor which can tolerate 5-watt dissipation with a typical heat sink in a 25-degree-centrigrade (C) ambient has about 1 watt collector dissipation. The dissipation caused by leakage at cutoff is lower, being in the range of 0.2 watt. In addition, a dissipation of about 0.5 watt is caused by the base drive. It appears that the maximum practical pulse-repetition rate is in the range where the transient dissipation approximates the maximum steady-state dissipation of, typically, 1.5 watts. Assuming in this example that P_T = 1.5 watts, P_L = 30 watts, T = 15μsec, equation 2 gives a practical pulse-repetition rate of 10,000 pulses per second. Since in this circuitry two pulses occur per cycle, this value corresponds to magnetic-amplifier operating frequency of 5 kc. In practice, the allowable transient dissipation is, of course, affected by such factors as the required output power, available heat sink, and specified ambient temperature range. With presently available power transistors, however, switching rates in the range from 1 kc to 10 kc seem most advantageous.

The switching time of a power transistor is much less in the common-base than in the common-emitter connection. The power gain is also greatly lowered because the current gain in common-base connection is less than unity. The power-handling capability for a given internal dissipation is also significantly reduced.[1] However, in applications where reduced gain and output power can be tolerated, the switching rates indicated above can be increased by at least a factor of 10, using the common-base connection. Magnetic amplifiers can be successfully operated in the megacycle range, and so the speed of the circuit in power applications is currently limited by the switching time of the transistors.

To yield maximum figure of merit, a high-frequency magnetic amplifier should be designed so that the saturated (air-core) reactance of the magnetic core approaches the value of the load resistance.[7] If this design requirement is met, the magnetic amplifier will not have a sharp output pulse because the air-core inductance limits the rise time when the amplifier fires, even though the magnetic material saturates completely. In order to obtain a rectangular output pulse (short-rise time) the magnetic stage should be designed so that the time constant formed by the ratio of saturated inductance to load resistance (L_s/R_L) is slight compared to the transistor switching time. As a result, the maximum figure of merit cannot be realized if the transistor dissipation is minimized. The net effect is that the theoretical power gains previously indicated for a magnetic amplifier are somewhat reduced in practice. This design compromise allows a very good switching pulse to be delivered to the transistors.

Experimental Example

The circuit of Fig. 5 has been constructed, and uses two Minneapolis Honeywell *H-2* transistors in the output stage. The transistor inverter supplies the magnetic amplifier with 9 volts at 5 kc. Germanium diodes are used for the self-saturating rectifiers, and to provide the base bias. The magnetic-amplifier reactors consist of 650 load turns and 800 control turns wound on 1/2-in. (inch) inside-diameter, ceramic bobbins having 32 wraps of 1/2-mil *4-79* Mo-Permalloy. The inverter also uses *H-2* transistors, but lower power transistors are suitable.

A control characteristic of the amplifier is shown in Fig. 6. With a 50-volt supply, the maximum output power is 47 watts. With only the winding resistance in the control circuit (52 ohms) the amplifier has a total power gain of 15 × 10^6 (72 decibels) and response time of 85 milliseconds. The response time can be decreased to a minimum value of 0.1 millisecond by increasing the control-circuit resistance and suffering a corresponding decrease of power gain. A delay of 1 cycle (0.2 millisecond), a voltage gain of 9, and a power gain of 32,000 (45 decibels) will result with a control-circuit resistance of about 20,000 ohms. By a proper selection of the number of control turns, the input impedance, and

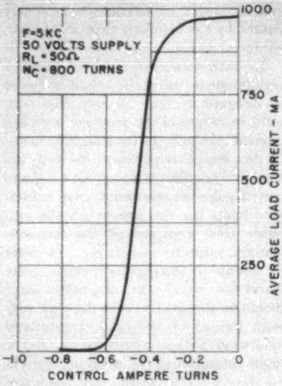


Fig. 6 (left). Control characteristic of magnetic-transistor amplifier

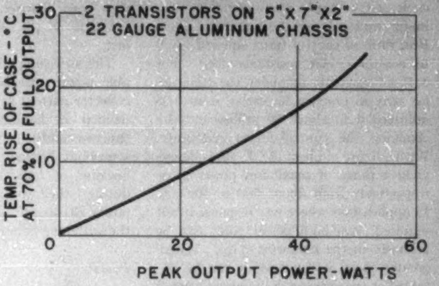


Fig. 8 (right). Transistor temperature rise at pulse repetition rate of 10,000 pulses per second

control-voltage sensitivity, for a given power gain and delay, can be greatly altered.

Fig. 7 shows the output wave form of the inverter and the output wave form of both the magnetic and the transistor stage operating at about 70% conduction with a peak output of 47 watts. This operating point is near the condition of maximum transistor dissipation. The power dissipation per transistor is approximately 5 watts. With the two transistors mounted on a 5-in. by 7-in. by 2-in. 22-gauge, aluminum chassis, and with a mica insulating washer, the temperature rise of the mounting base is 24C. With 5-watt internal dissipation in the *H-2* transistors, a maximum mounting base temperature of 70 C is permissible as has been indicated by the manufacturer. A maximum ambient temperature of 46 C is, therefore, tolerable with this power output. Fig. 8 shows the mounting base temperature rise at 70% conduction for various values of peak output power. It will be noted that the temperature rise is disproportionately high at high output power. The increased rate apparently results from the excess dissipation caused by increased leakage at higher junction temperatures, and by increased switching time at high values of collector current.

Circuit Variations

Variations in the basic circuit shown in Fig. 5 are possible. Fig. 9 shows a doubler-type magnetic amplifier used in a transformer-coupled version of the circuitry. Although it requires an additional transformer, this circuit allows the magnetic amplifier to be operated at a more advantageous voltage level. Using a higher supply voltage, the power-handling capability of the germanium diodes is increased, and their forward voltage drop is of less significance. Also the use of a doubler circuit is advantageous because it does not place full supply voltage on the rectifiers. This configuration is preferable where the maximum possible power is required from the magnetic stage. Unlike the circuit of Fig. 5, the transistors operate on alternate half-cycles.

Fig. 10 shows a circuit which amplifies a-c signals. The transistors are biased off when there is no input signal and operated class *B* into a center-tapped output transformer. The magnetic-stage pulse width modulates each transistor for $1\frac{1}{2}$ cycle of the signal frequency. The control windings are connected so that each side is turned on by opposite polarities of the a-c signal. The carrier-frequency pulses can be filtered, and the signal-frequency power is delivered through a transformer to the load. It is apparent that the a-c signal frequency must be considerably less than the carrier frequency, since each 1/2 cycle is formed from many carrier-frequency pulses. The single-ended circuits shown in Fig. 5 and 9 can also amplify a-c signals, if the d-c output is modulated around a quiescent output level and the signal-frequency power capacitively coupled to the load. A carrier-frequency a-c output can be delivered to the load if the output transistors in Fig. 9 are operated on alternate 1/2 cycles into a center-tapped transformer. The center-tapped circuits inherently place double voltage on the transistors, and the output power of these circuits is halved.

In many applications the carrier-frequency pulses in the output are not desirable. The output of these amplifiers can be filtered, and the input wave form thereby recovered and accurately reproduced. If reduced output power can be tolerated, only one transistor need be used in the output stage of most of these configurations. Greater signal sensitivity can be obtained by using either a low-level magnetic amplifier[11] or a transistor stage as a preamplifier.

It is apparent that it is possible to design a similar group of circuits using a

Fig. 7(A). Output of 5-kc inverter

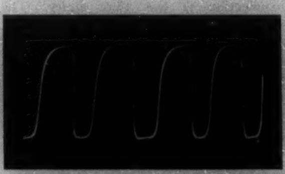


Fig. 7(B). Magnetic amplifier output pulses

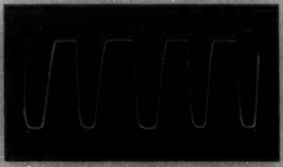


Fig. 7(C). Output of pulse-width-modulated power transistors

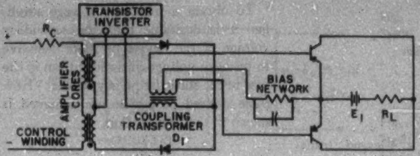


Fig. 9 (above). Transformer-coupled circuit

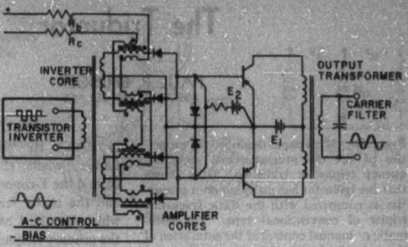


Fig. 10 (right). Switching transistor a-c amplifier

standard power-frequency magnetic stage for control of the transistors. Although the firing angle of a sinusoidally operated magnetic amplifier does not vary linearly with control, the average output can be controlled linearly if the direct current for the switching transistors consists of a rectified, but unfiltered, sine wave. It is believed that the following advantages will justify operation from a high-frequency square-wave source such as proposed in the foregoing, even though additional circuit components are required.

1. The power-handling capability of the transistors is approximately doubled, since they can operate at peak collector voltage throughout the cycle.
2. Much shorter response times are possible, high-gain a-c amplifiers can be built, and much smaller output filters result at the higher operating frequencies.
3. A higher figure of merit is obtained in the magnetic amplifier, and greater power gain is, therefore, possible with a given time delay.
4. The magnetic stage is smaller in size and weight, and is possibly less expensive.
5. The cost of additional inverter components (two low-power transistors and a core) is normally a small part of the total device cost.

Circuit Features and Applications

The circuit techniques proposed in this paper are most advantageous where an efficient linear amplifier of high output power is required. Transistors could be developed which would control many amperes and hundreds of watts of power when operated in the switched mode. As these power levels are realized, the inherent inefficiency and the resultant high internal dissipation make class *A* operation increasingly less practical. Consequently, the advantages of linear control of the average output by pulse-width modulation techniques are increasingly attractive. The amplifier-response time with pulse-width modulation is, of course, longer than with class *A* operation, being a minimum of 1/2 of the carrier frequency (typically 0.05 to 5.0 milliseconds). On the other hand, the relatively slow response is not disadvantageous in most of the probable applications: for example, where time delays from mechanical elements are present.

The linearity of a circuit employing pulse-width modulation is determined primarily by the linearity of the modulating device, and is largely independent of the parameters of the modulated element. A very linear and stable amplifier can be obtained if the gain of the magnetic stage is greatly degenerated by negative feedback. If the switching time of the transistors remains negligible, excellent linearity can be obtained at a high power level by amplification of the modulator output pulses with successive stages of transistors.

These circuits are generally applicable where a high-power high-gain moderate-speed linear amplifier, having the small size, ruggedness, and reliability of magnetic and semiconductor devices, is desired. Such amplifiers are commonly required in servo systems, for example, and in regulated power supplies. Future development of very-high-power transistors should make possible an excellent means of d-c motor-speed control and generator field control.

Conclusions

The power-handling capability and stability of a transistor amplifier can be greatly increased by the use of pulse-width modulation for proportionately varying the average output power. A high-frequency magnetic amplifier operating from a square-wave source provides an excellent means of pulse-width modulating of power transistors. The magnetic amplifier functions both as a stable pulse-width modulator and as a high-gain preamplifier. In addition, the magnetic stage is small and rugged, has speed commensurate with the maximum practical transistor-switching rates, and provides a means of matching input impedance to a wide range of control sources. The basic technique, which combines the high linearity of a magnetic stage with the high power-handling capability of switching transistors, can be applied to various circuit configurations. The resultant amplifiers feature small size, ruggedness, high gain, high power, moderate speed, and linear control of the average output.

References

1. Junction Transistors Used as Switches, R. L. Bright. *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 111–21.

2. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, G. H. Royer. *Ibid.*, July 1955, pp. 322–26.

3. Transistor Demodulator for High-Performance Magnetic Amplifiers in A-C Servo Applications, R. O. Decker. *Ibid.*, Mar. 1955, pp. 121–23.

4. Switching Transistors Used as a Substitute for Mechanical Low-Level Choppers, A. P. Kruper. *Ibid.*, pp. 141–44.

5. Transistor Power Amplifiers with Switched Mode of Operation, A. G. Milnes. *Ibid.*, vol. 75, pt. I, July 1956, pp. 368–72.

6. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.

7. High-Frequency Operation of Self-Saturating Magnetic Amplifiers, H. William Collins. *AIEE Transactions*, vol. 74, pt. I, Sept. 1955, pp. 500–05.

8. Operational Magnetic Amplifier with Audio-Frequency Transistor Power Supply, R. O. Decker, Francis Gourash. *Ibid.*, pp. 490–96.

9. Application of High Frequency Magnetic Amplifiers, H. William Collins. *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 11, 1955, pp. 136–44.

10. Magnetic Amplifier Circuits (book), William A. Geyger. McGraw-Hill Book Company, Inc., New York, N. Y., pp. 207–18.

11. Conference on Magnetic Amplifiers. *AIEE Technical Publication T86.* "Capacitively Coupled Magnetic Amplifiers," H. William Collins. Joint AIEE, Institute of Radio Engineers, New York, N. Y., Instrument Society of America, New York, N. Y., 1956, pp. 211–20.

The Triductor

PAUL P. BIRINGER
ASSOCIATE MEMBER AIEE

Synopsis: This paper describes the properties of the d-c premagnetized static-frequency tripler or triductor. It is shown that the triductor has many superior properties as compared with the static frequency tripler of conventional type. The automatic or manual control of the saturation of the individual iron cores by superimposed d-c magnetization makes a number of applications practical. Of particular interest are the uses of the triductor as a static frequency transformer with regulated output voltage and its application as a highly voltage-sensitive amplifier.

STATIC frequency changers, as such, have long been known. They came into early use in the field of radiotelegraphy. However unlike most other electric devices, they have not undergone any continuous development during the last few decades. The improvement of magnetic materials with low losses at high flux densities has facilitated the development of static frequency changers during recent years. Only the static frequency tripler and the d-c premagnetized static frequency tripler or triductor are considered in this paper.

Triductors for frequency changing purposes have been used for a few years in the U.S. and Canada. The installations consist of units having 50-kw maximum output power. The use of the triductor as a voltage-sensitive amplifier has been investigated at the University of Toronto and shows considerable promise.[1]

Properties of Static Frequency Tripler

The important characteristic of the static frequency tripler is that its operation is based on the nonlinearity of the **B-H** characteristic. The simplest connection is shown in Fig. 1. Assuming that the terminal voltage ($\mathcal{E}_1$) is so high that the transformer operates above the kneepoint of the **B-H** characteristic of the iron core, then terminal voltage, flux, and the secondary voltage will be distorted. Due to the symmetry of the **B-H** characteristic only odd harmonic components are obtained across the secondary terminals. Any of these harmonics may be filtered out and utilized on the secondary side of the transformer.

Another connection specially for frequency tripling was described by Spinelli in 1912.[2] He used a 4-legged transformer, where the three outside legs carried the *Y*-connected primary windings and the center leg a single winding, in which the third harmonic component of the primary frequency could be obtained; see Fig. 2.

The same result can be achieved by using three identical single-phase transformers with star-connected primary windings and series-connected secondaries, as shown in Fig. 3. The latter is the conventional static frequency tripler.[2-4] This connection has the following properties:

1. The magnitude of the secondary voltage is highly dependent on the value of the primary voltage. This property is not desirable if one wants to use the device as a frequency transformer.
2. The primary power factor is very low, because of the high magnetizing current required at the high flux densities.
3. The reactance behind the third harmonic secondary emf (electromotive force) is relatively high and it limits the high-frequency output power to a small value compared with the size of the transformer.
4. For 29-gauge silicon iron core and a pure resistive load, the transformer efficiency is relatively low, 70 to 80% (per cent), due to the high losses in the iron and in the primary windings.
5. The primary current is distorted. It contains a large percentage of fifth-harmonic component, and for this reason the primary power factor cannot be corrected by capacitors across the primary terminals.

To obtain practical frequency transformation, it is desired that the secondary voltage can be kept constant, independent of the primary voltage variations, that the primary power factor can be corrected, and that the dimensions of the transformer compare favorably with power transformers of the same output power.

To obtain a high gain voltage amplifier it is desirable that the secondary voltage of the device be more sensitive to primary voltage variations than is the case in the static frequency tripler. How these desired properties are achieved is shown in the next section.

Connection and Properties of Triductor

Fig. 4 shows the connection of the triductor designed for operation as a frequency transformer. The diagram shows two core elements in each a-c phase. They are provided with separate primary N_p, secondary N_s, d-c $N_{d\text{-}c}$, and compensating N_c windings. The general properties of this connection are described in the following:

1. The d-c premagnetization shifts the position of the flux wave on the flux axis. The amplitude of the induced third-harmonic emf depends on both the magnitude and position of the a-c flux wave. Consequently, the secondary voltage and power output of the triductor can be influenced by changes in either the d-c magnetization or in the primary voltage. The change in the secondary voltage due to primary voltage variations may be counteracted by a corresponding change of the d-c premagnetizing current. In this way the secondary voltage can be kept constant.
2. With the help of series C_2 or parallel C_3 condensers the high secondary impedance can be counteracted. The possible output power for the same core size will increase. This increase very often corresponds to two or three times the original output power without reaching the region of ferroresonance.
3. Because of the condensers the frequency-transformed power will increase, while the losses will increase only very slightly, therefore the efficiency will have a much higher value than before (90 to 97%).
4. The power-factor-correcting condensers C_1 are placed across a separate winding N_c where the induced emf is already distorted. This connection makes a correction to nearly unity power factor possible. Placing the condensers on the primary side would result in a correction of the fundamental frequency component of the magnetizing current only. At the high operating flux densities it was found that the fifth harmonic current component can reach 50% to 60% of the fundamental. Thus the connection shown in Fig. 4 was used with very good results.

The connection shown in Fig. 4 possesses some unique properties which are advantageous if the triductor is to be used as an amplifier. These properties are as follows:

1. If the primary voltage is kept constant and increasing d-c magnetization is applied to the triductor, the secondary emf dis-

---

Paper 56-715, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 26, 1956; made available for printing May 1, 1956.

Paul P. Biringer is with the University of Toronto, Toronto, Ont., Canada.

The research reported was supervised by Profs. G. F. Tracy and J. M. Ham, of the University of Toronto, and it was made possible through support extended to the Department of Electrical Engineering, University of Toronto, by the Defence Research Board of Canada, under grant no. *4001–02*, *Project 048-40-01-02*.

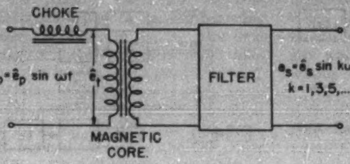


Fig. 1 (left). Static frequency multiplier

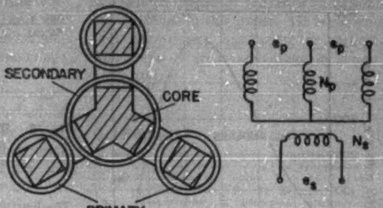


Fig. 2 (right). Static frequency tripler by Spinelli

appears at some definite d-c value. The transfer characteristic is shown in Fig. 5.

2. The triductor will be much more sensitive to primary voltage variations in some regions, than the static frequency tripler. Hence there will be a high gain relating the primary voltage variations to the secondary voltage variations or power output; see Fig. 6. External feedback can be applied to increase this gain.

3. The time delay (primary to secondary) can be reduced without significant changes in the gain. However, one must accept a somewhat lower efficiency.

4. For transients the device has a lead-network character.

These properties are of definite advantage if the triductor is used in a closed-loop regulating system.

Calculation of Triductor Characteristics

The basic equation of the static frequency tripler may be deduced from Kirchhoff's law and from the magnetomotive force law. It yields, for instantaneous values

$$\sum_{k=1}^{3} Hg(b_k+\beta)=\frac{3N_s i_s}{l} \tag{1}$$

where

H=resultant magnetizing force, instantaneous value
b=flux density of fundamental frequency, instantaneous value
β=high frequency flux density, instantaneous value
N_s=number of secondary turns
i_s=secondary current, instantaneous value
l=length of the mean magnetic path
k=number of the primary phases
g=function

The solution of equation 1 is obtained in the two special cases of open-circuited ($i_s=0$) and short-circuited secondary ($\beta=0$) by a trial-and-error method. For instance, in the case of open-circuited secondary the flux densities b_1, b_2, and b_3 are calculated for various time t_1, t_2, . . ., etc. It is known that the third, ninth, etc., components of the flux density are equal in each phase at every instant. Thus one must add the same but unknown flux density β to b_1, b_2, and b_3 so as to satisfy equation 1. Equation 1 is satisfied if the sum of the resultant magnetizing forces H is zero. Because the value of the resultant magnetizing force H is dependent the unknown sums $b_1\pm\beta$, $b_2\mp\beta$ and $b_3\pm\beta$, the result cannot be obtained directly. The use of a nonlinear nomogram can facilitate the solution.

The secondary voltage and current of the loaded static frequency tripler are calculated by solving the difference equation of the secondary side. The following equation is written if a load containing resistive R, capacitive C, and inductive L components in parallel is assumed

$$\sum_{k=1}^{3} M(\varphi_k+\phi)=-9N_s^2\times$$

$$\left[C\frac{\Delta^2\phi}{\Delta t^2}+\frac{1}{R}\,\frac{\Delta\phi}{\Delta t}+\frac{1}{L}\,\phi\right]10^{-8} \tag{2}$$

where

M=resultant magnetomotive force
φ=flux of the fundamental frequency
ϕ=high frequency flux

Equation 1 can be used with some modification for the determination of the secondary open-circuit voltage and the secondary short-circuit current of the triductor.

For any given value of the line-to-line voltage the fundamental components of the flux and of the flux density b across the two elements A and B of the triductor may be calculated. Similarly the flux and flux density (B_s) due to the d-c magnetizing current I can be obtained with the help of the **B-H** characteristic of elements A and B. These two elements are magnetized in opposite directions with direct current, and hence the sum of the flux densities due to I is zero in the two elements

$$B_{sA}-B_{sB}=0 \tag{3}$$

The value of a-c magnetizing force is equal in both elements A and B, at every instant, because the same current is flowing through the elements.

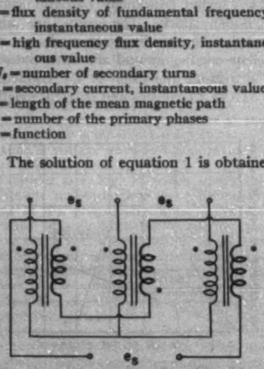


Fig. 3. Static frequency tripler

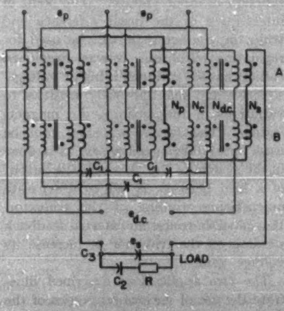


Fig. 4. The triductor with compensating condensers on primary C_1 and on secondary C_2 and C_3 side

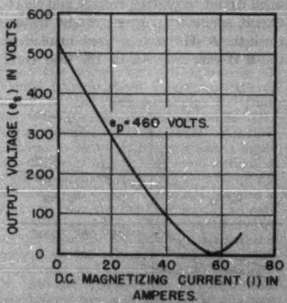


Fig. 5. Transfer characteristic of triductor. Secondary voltage versus d-c magnetization at constant primary supply voltage

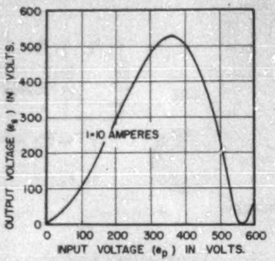


Fig. 6. Transfer characteristic of triductor. Secondary voltage versus primary supply voltage at constant d-c magnetization

$$h_{kA}=h_{kB}=h_k^* \tag{4}$$

The flux density at fundamental frequency (b_{kA}) in element A and the flux density (b_{kB}) in element B change from instant to instant so as to satisfy equation 4 and equation 5

$$b_{kA}+b_{kB}=b_k^* \tag{5}$$

Thus the value of a-c magnetizing force h_k^* can be found for any given line-to-line voltage and d-c magnetizing current I with the help of a $B'-H'$ characteristic, which is obtained in the following way

1. Two auxiliary curves for B_0+b_{1A} and $-B_0+b_{1B}$ are constructed, the former by shifting the original curve horizontally to the left a distance H_0 and the other by shifting the original curve to the right a distance H_0.
2. Curve $B'-H'$ is drawn, the ordinates of which are obtained by adding the corresponding ordinates of the curves B_0+b_{1A} and $-B_0+b_{1B}$; see Fig. 7.

Equation 1, of the static frequency tripler, can be used now to obtain the open-circuit secondary voltage and short-circuit secondary current of the triductor if the **B′-H′** curve is used instead of the **B-H** characteristic of the iron core.

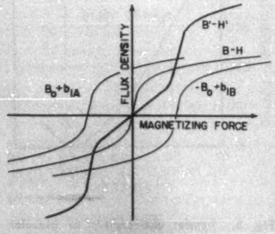


Fig. 7. The $B'-H'$ characteristic of two series-connected triductor elements

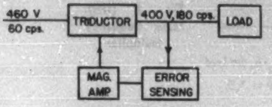


Fig. 8. Block diagram of triductor with regulated output voltage

The load voltage and load current are obtained by solving the difference equation 2 of the secondary side.

Applications of Triductor

Two applications are described here, the first being the use of the triductor as a frequency changer with regulated output voltage. The block diagram in Fig. 8 is related to an installation where several 50-kw triductors are used for induction heating purposes. The net d-c magnetizing current of the triductor is influenced by the deviation of the output voltage from the desired value. When the load conditions or the primary voltage changes, an error voltage is fed back to the triductor so as to restore the original output voltage. Provision is made for manual setting of the desired output voltage or output power. This power triductor has compensating capacitors on the primary side and parallel capacitors on the secondary side. The efficiency at full output, 50 kw, is 92%.

Another application is shown in Fig. 9, where the triductor is used as a voltage-sensitive amplifier in a voltage regulator for synchronous generators.[1] Fig. 10 shows how the output of the triductor is related to the 3-phase input voltage, when the d-c separate magnetization is constant. The same figure gives the output current of a linear element as a function of input voltage. The point of intersection of the linear reference and triductor characteristic can be used for regulating purposes. It was mentioned earlier that the triductor is highly voltage-sensitive and simultaneously sensitive to variations in the separate d-c magnetization. The latter property can be used for stabilizing purposes, as is done in the system of Fig. 9, where the output of a lead network is fed to a separate magnetizing winding of the triductor. It is possible to use an external feedback winding on the triductor to increase its voltage gain.

The two applications described illustrate the use of various properties of the triductor. There are other fields where, again, other properties are of advantage. For instance, in arc furnaces the high frequency helps to obtain a stable arc and the secondary load is symmetrically distributed on the primary side, whereas in practically all other methods the furnace constitutes a single-phase load.

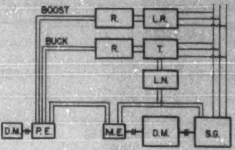


Fig. 9. Block diagram of voltage regulator for synchronous generators employing triductor. R=rectifier, LR=linear reference, T=triductor, LN=lead network, DM=driving motor, PE=pilot exciter, ME=main exciter, SG=synchronous generator

Design and Performance Data

Generally the design of d-c premagnetized frequency triplers is such that the iron cores are not saturated by the applied a-c voltage. The saturation and the generation of third harmonics is achieved by the superimposed d-c magnetization. This method has the disadvantage that only part of the core is participating in the a-c flux reversals. Without the application of any direct current the cores of the triductor are already saturated during a part of the cycle. The result of this design is a better utilization of the cores, which has a great importance in large units.

The core material of the 50-kw triductors is high-quality silicon steel. Each of the six units are made out of standard *E-I* laminations: *VEI* 250 ×4$^1/_2$ inches (Allegheny Ludlum) using a 6-inch stack. The cores are air-cooled. The windings are made out of hollow conductors and are water-cooled. The over-all dimension of the triductor is 32×15×12 inches and its total weight

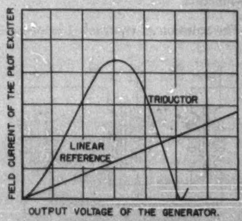


Fig. 10. Operating characteristics of regulator

Vol. 7
Pt. 1
1956

is 620 pounds. There are six condensers connected to the compensating winding, as shown in Fig. 4. The total rating of these condensers is 90 kilovars at 460 volts, 60 cycles per second. The rating of the condensers which are in parallel with the load is 225 kilovars at 180 cycles per second. This high value was required because of the low power factor of the load which is an induction heating coil. In the case of resistive load the secondary condenser requirement is about 30 to 45 kilovars.

The triductor described is able to hold the secondary voltage constant within ±5% if the primary voltage variation is less than ±15%. The power factor of the load must not be lagging at any point of the operation. The required d-c control power is 650 watts, which is not included in the calculation of the efficiency at full power output. There are several 50-kw triductors of similar design which have been in continuous operation since 1953.

The requirements on the voltage-sensitive triductor are such that the use of special core materials is necessary. Mumetal *UBF* stampings have been used in the experimental unit[1] and the dimensions of each element were $2^4/_8$ by $1^1/_8$ by 1 inch. The maximum output power was 18 watts. The response time was less than one cycle of the 60 cycle-per-second supply.

Summary

The properties, basic equations, and practical applications of the d-c premagnetized frequency tripler or triductor are described in this paper. It is shown that this device has many properties which are of great advantage in some applications.

References

1. The Voltage Stability of the Synchronous Generator With a Third Harmonic Type Voltage Regulator, **P. P. Biringer.** *Ph.D. Thesis*, Department of Electrical Engineering, University of Toronto, Toronto, Ont., Canada.

2. Transformatore Elettico Della Corente Alternata Trifasa In Correlli Monofasa Con Frequenza Triplicata, **Francesco Spinelli.** *Italian Patent No. 124.825.*

3. Magnetic Frequency Conversion, **L. C. Harriot.** *Proceedings*, National Electronics Conference, Chicago, Ill., vol. 9, 1953, pp. 78–87.

4. Magnetic Frequency Multipliers, **L. J. Johnson, S. E. Rauch.** *AIEE Transactions*, vol. 73, pt. I, Nov. 1954, pp. 448–52.

◆

Discussion

N. L. Kusters and **W. J. M. Moore** (National Research Council, Ottawa, Ont., Canada): By the addition of d-c premagnetization to the saturating transformer type of magnetic frequency tripler, Dr. Biringer has produced an effective means of controlling the third harmonic output of this type of device. The circuit makes use of two transformers in series in each phase, thus permitting a d-c magnetomotive force to be applied in opposite polarity to each of the two cores in each phase, as well as prohibiting the generation of any fundamental or third harmonic voltage in the d-c circuit. The composite magnetization curve for each phase produced by this type of connection has certain characteristics which are pertinent to this application and it is of interest to consider how these affect the third harmonic voltages induced in the secondary circuit. For the purpose of this discussion all harmonics greater than the third are neglected.

Fig. 11 shows an idealized magnetization curve for the condition of zero d-c premagnetization. If the secondary is open-circuited, a sinusoidal magnetomotive force of fundamental frequency only will exist in each phase and it will be noted that this produces a square-shaped flux wave in the cores. If this is differentiated with respect to time, a voltage wave with a pronounced third harmonic peak is revealed.

Consider now Fig. 12 which shows the composite magnetization curve for conditions of d-c premagnetization. Note the presence of an additional "break" in the curve which is in the opposite sense to that resulting from magnetic saturation. Applying a sinusoidal magnetomotive force as before, it is seen that this condition yields a peaked flux wave and thence a double-peaked voltage wave.

It is apparent from the two figures that the two types of break in the composite magnetization curve produce third harmonic voltages which are 180 degrees out of phase with each other. This explains the form of the control characteristics shown by the author in Figs. 5 and 6. In connection with these figures, it might be appropriate to mention that the output voltages to either side of the zero output points shown should indicate a phase reversal has occurred.

All this, however, does not explain how power is transferred from the primary at one frequency to the secondary at another frequency, as the previous remarks have dealt with magnetizing currents only, and these in the ideal case convey no power. The author, in his paper, presents a method which relates the input to output wave forms for the case of power transfer. This analysis is valid for any shape of magnetization curve but involves a tedious step-by-step procedure which obscures the process by which energy is transferred. If a more idealized magnetization curve is assumed, it is possible to use the "gating" concept commonly used in magnetic amplifier analysis. This method clarifies the process of energy transfer, at least in a qualitative way. It was used previously by Johnson and Rauch (reference 4 of the paper) in the analysis of a somewhat different third harmonic generator in which the load was connected between the neutral of the supply and the neutral of three saturating reactors. In that case, power was switched into the load by the cores reaching saturation at a certain instant in each half-cycle of voltage in each phase, so that the sequence of pulses received by the load formed a triple frequency current. From the primary viewpoint, however, the power delivered by each phase was at fundamental frequency.

In the circuit proposed by Dr. Biringer, the third harmonic load is coupled in by an equivalent delta connection. The process of energy transfer remains basically the same but the explanation is somewhat different. Before saturation occurs, the total voltage induced in the secondary winding is zero. As soon as one transformer saturates, this voltage is no longer zero but equal to $\sqrt{3}$ times the phase voltage which existed across the saturated transformer before saturation took place.

A few of the points raised in the paper provoke some thought. In particular, it is interesting to read that more effective power factor correction is obtained by means of a separate secondary winding. Possibly this is due to the nonlinear coupling which exists between the primary and secondary, and other windings of this device. Perhaps the author could expand on this point.

The statement that the device exhibits a lead network characteristic is also an interesting one. If the input in this case is considered to be the primary voltage of fundamental frequency, and the output the secondary voltage of harmonic frequency, this point is rather difficult to see and it would be appreciated if some clarification could be provided.

Andrew A. Halacsy (Jeffries Transformer Company, Hollywood, Calif.): This paper is an interesting continuation of Dr. Biringer's work on frequency doublers presented last winter and it may be linked with Conference Paper No. 56-229 presented last winter by W. McMurray of General Electric about frequency triplers, similar to Dr. Biringer's Fig. 3. That one was parallel connected while Dr. Biringer's primary is in series. Another difference is that Dr. Biringer uses a separate d-c winding. I was very interested to see the compensation for power factor by an inductively coupled capacitor because a power factor correction enables the adjustment of the power factor not only by changing the capacitor but also by changing the tapping on the respective winding in the transformer, as I used it some 20 years ago. Dr. Biringer compensates for high load impedances, too, and by this means any load, not only a resistive load, can be used. Could a capacitive load be compensated the same way? The following few points I should like to have clarified.

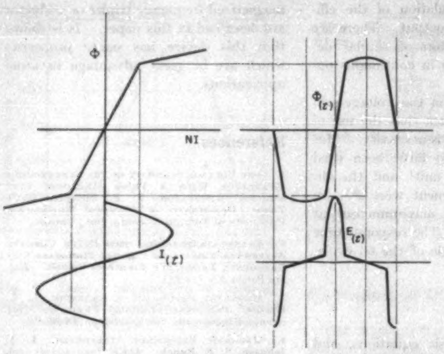


Fig. 11. Idealized magnetization curve for condition of zero d-c premagnetization

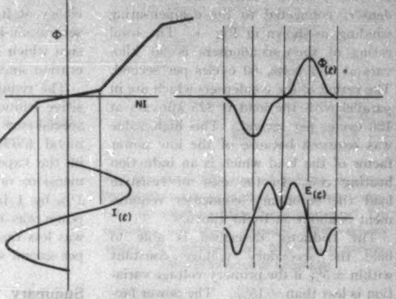


Fig. 12. Composite magnetization curve for conditions of d-c premagnetization

It is mentioned that the triductor will be much more sensitive to primary voltage variations in some regions than the static frequency tripler. Is this not in contradiction with the requirements mentioned in point 1 in the section "Connection and Properties of Triductor?"

Equations 3 and 5 are true if they are transformed into voltage equations. But, as far as flux densities are concerned, one can speak only of imagined flux densities because the resultant flux densities will appear nowhere. Equation 4 is true, of course.

The method of typing is confusing in equations 1 and 2, where a function appears under the sign of summation. Many may assume that the expressions $H(b_k+B)$ and $M(\phi_k+\phi)$ represent products and not a function.

P. P. Biringer: The author wishes to thank the discussers for their interest. The points they have brought out present an opportunity to give the author's views on these matters.

In Mr. Kusters' and Dr. Moore's discussion, the first question concerns the method of analysis. It is quite true that the "gating" concept clarifies the process of energy transfer in a qualitative way. The trouble is that the results of this type of analysis are based on an ideal **B-H** characteristic and if the **B-H** characteristic deviates from the assumed ideal one the results are useless for quantitative calculations. The author's intention is to present a method which can be used for design of triductors. It is appreciated that the step-by-step procedure for calculating the load current of the triductor is tedious, but the method is easily adapted for digital computers. In many cases, it is sufficient to know the no-load secondary voltage and the short-circuit secondary current of the triductor. In such cases equation 1 of the paper is simplified, and it is easily transformed to a nomogram which gives directly the high-frequency flux density or the high-frequency magnetizing force. Fig. 13 shows a nomogram for obtaining the high-frequency flux density, point by point. A further simplification results if it is assumed that the third harmonic component is dominant on the secondary side. In this case the maximum value of the high-frequency flux density and the maximum value of the secondary no-load voltage is obtained in a single reading on the nomogram.

The second question is in regard to the method of power factor correction. The necessity for a separate compensating winding may be illustrated as follows: Under normal operating conditions each core will be saturated during a part of the cycle. This results in a large and distorted magnetizing current, which causes a low power factor under load conditions also. The distortion of the magnetizing current is due not to the third harmonic current components, but rather to the fifth and higher harmonics. From this it follows that the power factor correcting condensers must be placed in such a way that they deliver current of fundamental, fifth harmonic and higher harmonic frequencies. It is important that the voltage across the condensers does not contain any third harmonic component. This is necessary to prevent the dissipation of the third harmonic power into this separate circuit. These requirements are met by using a separate compensating winding as shown in Fig. 4.

The third question brings out a very good point and it is difficult to give a short answer to this question. The transfer equation of the triductor for a sudden change in supply voltage was developed in reference 1. This theory and the actual measurements seem to justify the statement in the paper regarding the lead network properties. It is felt that the deduction of the relevant equations is too lengthy to be given here.

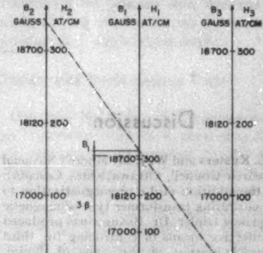


Fig. 13. Nomogram to obtain the high-frequency magnetizing force (β) of a static frequency tripler

Dr. Halacsy questions the possibility of compensating a capacitive load. It should be understood that condensers C_1 in Fig. 4 are compensating only the magnetizing current. Condensers C_2 and C_3 in the same figure are the ones which are used for compensating the reactance of the transformer and the inductive load. Naturally, if the load is capacitive condensers C_2 and C_3 can be reduced in size or entirely omitted. There is a limit to the capacitive load, namely, the point where resonance conditions will arise.

The sensitivity of the secondary voltage of the triductor to primary voltage variations is not of disadvantage because the secondary voltage can still be regulated by the d-c magnetization. This does not apply for the static frequency tripler; the secondary voltage follows the variations of the primary voltage and there is no way to prevent this. It is recognized, however, that for some applications, it would be better to have a constant secondary voltage which is not dependent on the primary voltage variations.

Dr. Halacsy is right that equations 3 and 5 represent imaginary flux densities and I am grateful for his suggestions.

Magnetic Logic Circuit Control System Design Considerations

R. I. VAN NICE
ASSOCIATE MEMBER AIEE

Synopsis: The design of switching circuits for industrial control has been developed chiefly in terms of relays. Magnetic logic circuits, which are switching elements based on the single-core magnetic amplifier, have properties different from relays and therefore require somewhat different design methods. One possible form of switching algebra for magnetic logic circuits is suggested and simple examples of its use are given. The steps in designing a complete system are discussed. The advantages of performing the important step of design verification on a simulator are pointed out.

THE SUBJECT of this paper is the design of switching control systems using as basic components a compatible set of magnetic logic circuits.[1] To clarify the meaning of the terms just used for those who may be unfamiliar with this subject, some background information and references[2] will first be given.

At one time, the term "switching circuit" was applied exclusively to a circuit containing moving-contact switches, such as manual or limit switches, relays, commutators, and stepping switches. With the development of vacuum-tube circuitry, particularly multivibrators, and of gas tubes such as thyratrons, it became evident that devices not containing moving contacts could perform switching functions. The basic characteristic of a simple switch is that it is either open or closed, so that its resistance to current flow is either high or low; the voltage across its contacts or the current through them is either essentially zero or not zero. It is this two-valued, or binary, nature of the quantities observed which is now used as the defining characteristic of a switching element. The list of switching elements now includes vacuum tubes and gas tubes, transistors, square-loop magnetic cores and ferroelectric capacitors, and others.

On analysis, it can be seen that the primary function of switching circuits is to perform logical operations. The variables of elementary logic are two-valued, namely, "true" and "false"; therefore they may be represented by one of the parameters of a switching circuit. The various connections of switching elements then yield the elementary logical functions of these variables such as AND, OR and NOT; and the mathematics of switching circuits is the algebra of logic, or Boolean algebra.[3,4]

Industrial control systems employing switching circuitry could be called logical computers. Input data are derived from push buttons, limit switches, friction switches, etc.; the logical computer takes the input data and computes the output data; and the output data operate solenoids, contactors, etc., which actuate the controlled equipment. Consider as an example an automatic elevator system with inputs from push buttons in the corridors, limit switches at the top and bottom of the shaft, and in the doors and gates, etc. From the state of these switches, the elevator control system must compute which direction the car is to go, at what floor it is to stop, when the door should open, etc.

Numerical computation can also be performed by switching circuits, since arithmetic operations are reducible to the more fundamental logical operations.[5] The history of digital computers shows that the first switching elements used were relays, which were then succeeded by vacuum tubes as higher speeds were demanded. Most recently, transistors and magnetic cores[6,7] have received much attention because they promise high reliability as well as high speed. The switching elements in logical computers have, on the other hand, undergone no change in kind until recently; relays are fast enough for most industrial control applications. The improvement of reliability is, however, of great importance, and for this reason transistor or magnetic core circuits are being proposed to replace relays.

A magnetic logic circuit is not an analogue of a relay, but is instead a device which performs a logical operation on a number of input variables. Thus, a relay network can be replaced by an interconnection of logic circuits which are logically equivalent; but the two networks may be quite different in appearance.[8]

The analysis or synthesis of switching circuits is usually carried out in terms of some form of Boolean algebra. Various modifications have been proposed for special purposes, and termed switching algebra,[2] computer algebra,[9] etc. It will be assumed here that the reader is familiar with the rudiments of Boolean algebra.

Mathematical Description of Magnetic Logic Circuits

The magnetic logic circuits discussed here are essentially magnetic amplifiers used as on-off elements. They are based on the single-core flux-reset amplifier described by Ramey and suggested by him as a computer element.[10] Some of the possible forms taken by these circuits are described in the literature;[1] it is not, however, necessary to have complete knowledge of their physical characteristics in order to design switching circuits. This section presents a mathematical description of the pertinent characteristics of the logic circuits, i.e., their logical, delay, and compatibility properties.

PHYSICAL, LOGICAL, AND DELAY PROPERTIES

A variety of names is given to the two values of the variables of logic and switching circuits. These include: True and False, Yes and No, Active and Inactive, On and Off, and 1 and 0. The last two sets will be used interchangeably here.

Physically, the voltage levels in the magnetic logic circuits represent the logical variables in the sense illustrated by Fig. 1. The transfer characteristic of the simple Ramey amplifier is given in Fig. 1(A) in a somewhat idealized form. The high output level is called ON or 1, and the low output is called OFF or 0. (The low output level is slightly negative because of the use of a biased rectifier across the output terminals to provide a path for magnetizing current.[1,10]) Intermediate output levels are not used to represent information and ideally to not occur in a computing system. Due to the choice of a voltage gain greater than unity, it is possible to obtain unambiguous operation; i.e., full output voltage from an element is more than sufficient to drive the next element to full output. Also, due to the nonlinearity of the transfer curve, the output is insensitive to variations in the input in both the 1 and the 0 ranges.

Another way of displaying these facts is

Paper 56-727, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 28, 1956; made available for printing April 30, 1956.

R. I. VAN NICE is with the Westinghouse Electric Corporation, East Pittsburgh, Pa.

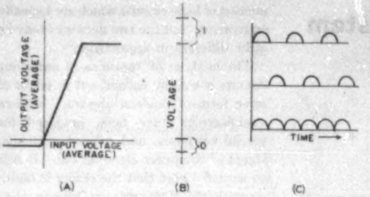
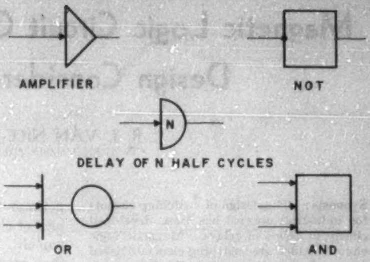


Fig. 1 (left). A—Transfer curve of Ramey amplifier. B—Tolerances. C—Three representations of "1"

Fig. 2 (right). Block diagram symbols for logic circuits



shown in Fig. 1(B). Any two voltages within a tolerance band about a design center are logically equivalent. These tolerances are valuable because they allow perfectly accurate operation in spite of the unavoidable variations in electrical parameters due to individual differences, aging, environmental conditions, and random fluctuations.

The transfer characteristic of Fig. 1(A) is given in terms of average voltages. On an instantaneous basis, the signals may be either half-wave or full-wave rectified sinusoids, approximately as shown in Fig. 1(C). Due to the half-cycle delay characteristic of the flux-reset amplifier, two phases of signals are used, which are logically equivalent but, of course, not physically equivalent. Similarly, it is possible to build magnetic logic circuits to give full-wave signals (e.g., the flip-flops described by Ramey[10]), and here again 1 and 0 are represented by full and zero output. (Systems in which 1 and 0 are represented by the two phases of signals will not be discussed in this paper.)

The input signals to a logical computer are obtained from switches which are not in general operated in synchronism with the computer power supply. If either rectified alternating current or pure direct current is used to supply the input signals, it is probable then that the first half-cycle of the input signal will be incomplete. As a result, ambiguous signals of value intermediate between 1 and 0 will exist in the system during transients. To secure correct operation, it is essential that the inputs be of more than 2 cycles in length so that at least one correct output may be obtained from the logic circuits fed by the inputs. It is necessary in uncertain cases to study the action of a system carefully to verify its correct operation when the inputs are asynchronously switched. Note that if magnetic logic circuits are to replace relays successfully, the time for 2 cycles must be comparable to the operate time for a relay.

Most numerical computers using magnetic logic circuits would use signals one cycle in length and synchronized with the power supply (clock). Extra equipment is required to convert input signals to this form, but the increase in speed may counterbalance the extra cost. Another factor is that numerical computers tend to have fewer inputs than the logical computers.

We turn now to the question of selecting a suitable mathematical symbolism for magnetic logic circuits from the variety of symbols which have appeared in the literature. (Neither the symbolism nor the algebra that follows is claimed to be novel.) The three significant aspects of a signal in such a system are magnitude, phase, and delay. The magnitude will be represented by a letter such as $A, B, C \ldots$ for inputs, and X, Y, Z for outputs; each of these is to be considered a binary variable which may take on only the values 0 and 1. Where it is desired to indicate phase, the subscripts a and b will be used. The operator D will be used to represent a 1/2 cycle delay in its operand; D^n represents n half-cycles of delay.

The simple Ramey amplifier which has been mentioned briefly introduces a half-cycle delay. If the input is denoted by A and the output by X, then

$$X = DA \tag{1}$$

Equation 1 implies a difference of phase between X and A; the two possible kinds of amplifier are more completely described by

$$X_a = DA_b \tag{2}$$

and

$$X_b = DA_a \tag{3}$$

Equation 2 expresses the fact that the amplifier accepts signals only of phase B, and gives an output of phase A (due to the phase of its power supply). The value of X equals the value of A during the preceding half-cycle (0 or 1).

The logical function variously called negation, NOT, inhibition, or complementation can also be performed by a single-core magnetic amplifier. A bar will be used to indicate the NOT function; as $\overline{0}=1$, $\overline{1}=0$. The equations for the two phases of NOT circuits can thus be written as

$$X_a = D\overline{A_b} \tag{4}$$

and

$$X_b = D\overline{A_a} \tag{5}$$

To illustrate the algebra of the symbols introduced so far, consider the cascade connection of two amplifiers described by equations 2 and 3, with $A_a = X_a$

$$X_b = DX_a = D(DA_b) = D^2A_b \tag{6}$$

Similarly, cascaded NOT circuits (using equations 4 and 5) give

$$X_b = D\overline{X_a} = D(\overline{D\overline{A_b}}) = D^2A_b \tag{7}$$

Therefore the result is the same in both cases: the output equals the input, delayed by one full cycle.

The logical function AND (coincidence) can be performed by an amplifying circuit described by Ramey.[10] The output of an AND circuit is one if and only if all the inputs are one. Algebraically, the AND function is equivalent to multiplication, and will be symbolized in the same ways. For a 3-input AND circuit, phases not indicated,

$$X = D(A \times B \times C) \tag{8}$$

Finally, the logical function of disjunction, or inclusive OR, is best performed by a simple semiconductor diode mixing circuit. The output of the OR circuit is zero if and only if all the inputs are zero. The plus sign will be used to indicate OR, but it must be kept in mind that simple addition is not meant. Because of the absence of an amplifier stage, no delay is introduced by an OR circuit. For three inputs equation 9 can be written [redacted]

$$X = A + B + C \tag{9}$$

The delays introduced by magnetic

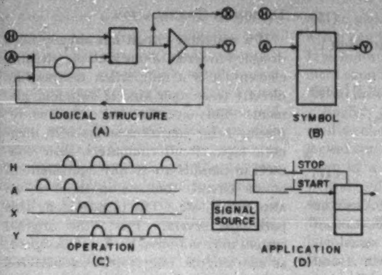


Fig. 3 (left). Flip-flop with hold input

A—Logical structure
B—Symbol
C—Operation
D—Application

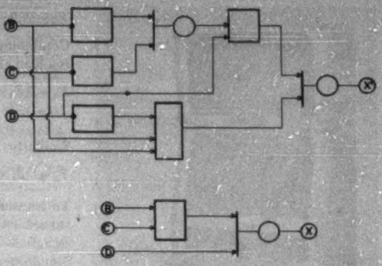


Fig. 5 (right). Block diagram

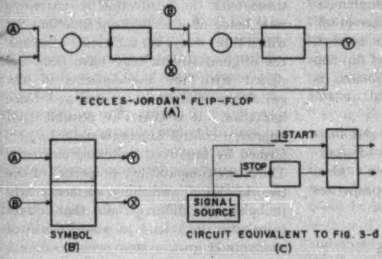


Fig. 4 (left). A—"Eccles-Jordan" flip-flop. B—Symbol. C—Circuit equivalent to Fig. 3(D)

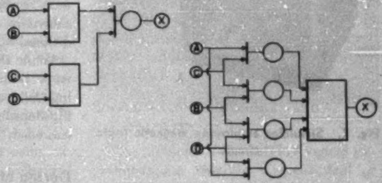


Fig. 6 (right). Logically equivalent circuits

logic circuits are in one sense unavoidable, since they are a result of the mode of operation of magnetic amplifiers. One consequence of these unavoidable delays is that extra amplifiers must occasionally be introduced in a system solely for the purpose of correcting phase or equalizing delays. Another consequence in an industrial control system employing a long chain of logical operations between input and output is that the total delay introduced might be excessive; this situation was not encountered in a study of a number of systems, assuming the use of 60-cycles-per-second magnetic logic circuits.

In many control systems, intentional delays are introduced as safety features, to provide time for an operator action, etc. A relatively short delay may be obtained by cascading several amplifiers, as indicated by equation 6. Longer delays are obtainable from a modification of the single-core amplifier[1]; for example, with a delay core and a buffer amplifier it is possible to construct a device whose equation is

$$X = D^{20}A \qquad (10)$$

i.e., a 10-cycle delay.

It is convenient to have a set of block diagram symbols for the magnetic logic circuits; one proposed set is shown in Fig. 2.

Compatibility Properties

A completely compatible system of logical elements could be interconnected in a completely arbitrary network. No physical elements are capable of arbitrary interconnection, and therefore one must specify the limits on their compatibility. For magnetic logic circuits, these limits include:

1. Phase: Inputs to an element must have the proper phase; correction is always possible by means of inserting an amplifier stage.
2. Delay: Same comment.
3. Loading: The number of inputs which can be fed from the output of a logic circuit is limited. The system designer must know how many inputs of each kind (AND, NOT) can be supplied by the outputs of each type of element. Of course, amplifiers can always be added so as to supply any number of inputs from one source.
4. Voltage drop: The OR element introduces a voltage drop, and therefore the number of OR circuits which can be cascaded is limited. Again, this limitation can be removed by the use of amplifiers.

Thus, if sufficient additional amplifiers are employed, magnetic logic circuits can be made completely compatible.

Storage of Information: Flip-Flops

Storage or memory is a vital function in any switching control system. One of the simplest uses is to store the fact that the START push button for an action has been operated more recently than the STOP push button, or vice versa. Storage is the distinguishing feature of a sequential switching circuit, whose outputs depend on the history of the inputs. Magnetic logic circuits are well suited to store signals since they combine the two properties necessary for regenerative storage: delay and amplification. The term "flip-flop" is applied to a number of bistable storage elements, two of which will now be described. Either of these could be used as a standard-package storage element.

The flip-flop with HOLD input may be built up from simpler logic circuits as shown in Fig. 3(A). Two outputs *X* and *Y* and two external inputs *A* and *H* can be noted in the circuit and, in addition, *Y* is fed back. The analysis of this simple logical feedback system is as follows:

$$X_b = D[(A_a + Y_a)\times H_a] \qquad (11)$$

$$Y_a = DX_b \qquad (12)$$

By substitution

$$X_b = D[(A_a + DX_b)\times H_a] \qquad (13)$$

The meaning of equation 13 can be put into words: "The output *X* is on during a *b*-phase half-cycle if *H* was on during the previous half-cycle and if either *A* was

Fig. 7. Simulator employing magnetic logic circuits

on the previous half-cycle or X was on one cycle ago (or both)." Admittedly, this sentence is awkward, and less useful than equation 13, but both give the essential operation of the circuit. Fig. 3(C) is designed to illustrate the operation more clearly. Starting with $X=0$ and $H=0$, $A=1$ does not turn the flip-flop ON; for $H=1$, $A=1$, the flip-flop does turn ON and stays ON, with the feedback signal Y replacing the momentary signal A; finally, the flip-flop is turned off by momentary loss of H. The input H is called a HOLD input, since its presence is necessary to hold the flip-flop ON. A block diagram symbol is shown on Fig. 3(B).

A typical use for the form of flip-flop just described is in connection with the START and STOP push bottons mentioned earlier [see Fig. 3(D)]. As a safety feature, it is commonly required that the STOP button interrupt a circuit, so that failure of that circuit due to any circumstance will cause the apparatus to stop. Therefore, the STOP button interrupts the HOLD input, while the START button applies the A input.

A second kind of flip-flop[5] is shown in Fig. 4(A). It is in some respects analogous to an Eccles-Jordan circuit. There are two inputs A and B and two outputs X and Y. The equations are

$$X_b=\overline{D(A_a+Y_a)} \qquad (14)$$

$$Y_a=\overline{D(B_b+X_b)} \qquad (15)$$

Combining, and using a familiar result in Boolean algebra

$$X_b=D\bar{A}_a\times\bar{Y}_a=D\{\bar{A}_a\times[D(B_b+X_b)]\} \qquad (16)$$
$$=D\bar{A}_a\times D^2(B_b+X_b)$$

Similarly

$$Y_a=D\bar{B}_b\times D^2(A_a+Y_a) \qquad (17)$$

To translate into words, output X can be turned on by momentary application of signal B and turned off by signal A. Simultaneous application of both A and B turns both outputs off. At all other times, outputs X and Y are complementary; i.e., if one is on, the other is off. Fig. 4(B) gives a block diagram symbol.

The use of this second type of flip-flop with START and STOP push buttons is shown in Fig. 4(B). Note that an additional NOT circuit is required.

While the flip-flops described give half-wave outputs, several methods of securing full-wave output are known. Also, flip-flops having other logical properties can easily be designed.

Design of Logical Computers

The goal of a design method is to provide a straightforward path from problem statement to practical solution. Part of the problem is usually to find either the most economical or the most reliable solution; also, one is restricted by limitations on the compatibility of the logical elements to be used. To one familiar with the difficulties of the design of switching circuits, it is not surprising that no straightforward design method is yet available. There are, however, a number of techniques which are of value to the designer in that they lead him to a number of near-minimal solutions, which can then be examined for economy, reliability, and compatibility.

Since this paper is not intended to be an exhaustive treatise on switching circuit design, the discussion will be confined to general considerations and to problems arising specifically from the use of magnetic logic circuits. References 2 and 11 through 16 may be consulted for further details on design methods. In the practical situation, one may be fortunate enough to be faced with so simple a system that the best solutions can be seen after only a small amount of trial and error. Or one may not be required to find the best solution, but only an adequate solution. At the worst, confronted with a truly complex system, the results obtained will depend on the time available and on ability or experience.

DESIGN BY TRANSLATION

If a switching circuit has already been designed in terms of one kind of switching element, it is a temptation to translate directly to another kind of switching element. This situation has arisen in replacing relay control systems with magnetic logic circuit computers, since contacts in parallel are simply equivalent to an OR circuit, contacts in series to an AND circuit, etc. The pitfalls along this path are several. First, the original system may not have been well designed; as an example, one system encountered had been modified piecemeal several times with the result that seven relays were being used to perform functions for which three relays are sufficient. Second, the original system may have been designed with the inadequacies of the switching elements to be used as a basic limitation; it is therefore possible that improved control functions could be performed by improved switching elements. Third, the compatibility properties of the two kinds of switching elements will probably be different, so that direct translation may lead to wasteful use of elements or to a system which will not operate properly. To avoid these pitfalls, it would seem best to take a fundamental approach to the problem of system design and to use previously designed systems mostly for comparison and checking. These comments apply particularly to the replacement of relays by magnetic logic circuits, and are intended as a warning to relay circuit designers who, because of years of experience, habitually think in terms of relays.

FUNDAMENTAL APPROACH

If one abandons the approach of direct translation, he may then follow steps such as these:

1. **Statement of the problem, including a complete description of the desired input-output relations and of requirements such as economy or reliability.**
2. **Development of near-minimal switching systems satisfying the input-output relations.**
3. **Examination of these possible solutions to determine which one best fits the remaining requirements of the problem statement.**
4. **Verification of the operation of the system by calculation or the use of a simulator.**

Each of these steps will now be discussed further.

INPUT-OUTPUT RELATIONS

The description of a logical computer in terms of what outputs are to result from given input conditions would seem at first glance to present little difficulty. Actually, an intimate familiarity with the apparatus to be controlled is essential. For example, considerable simplification can often result from a knowledge of what combinations or sequences of inputs will never occur.[14] Again, several possible logical patterns for control may exist, as in automatic elevators, and the designer must be able to determine somehow whether one pattern is mandatory or whether several are equally desirable. Finally, the designer must know the requirements of safety codes as they apply to the nature of the inputs which can be used and to provisions for fail-safe operation. Perhaps it would be ideal for this intimate familiarity with the equipment to be combined with expertness in switching circuit design in a single person, the designer. Practically, collaboration between two or more people is usually required, and is actually preferable because it leads naturally to a reanalysis of the control requirements for a system.

The first step in the specification of input-output relations is the definition of symbols for the input and output quantities. An important part of the definition in each case is a clear statement of when the quantity is 1 and when it is 0. For example, the actuation of a STOP push button S interrupts a circuit; is $S=1$ to mean that the button is pushed, or that the circuit is transmitting a signal? Either choice is permissible so long as there is no confusion in the mind of the designer. It is often convenient to define additional intermediate variables to represent those combinations of input variables which occur most frequently.

Switching circuits not involving the storage of information are called combinational circuits. Specification of input-output relations is relatively simple when the number of variables is small; either a table of combinations or a Karnaugh map[15] may be used, for example.

The requirement of information storage in a system makes it a sequential circuit, and complicates its specification. The flow chart introduced by Huffman[14] is probably the most complete method of representing the action of a sequential circuit. Again, the number of input variables which can be handled readily is limited to perhaps five. Other representations of sequential circuit operation are of course possible.[16]

Methods which have been suggested for handling larger numbers of variables include three and higher dimensional representations.[15] More obvious expedients are to split a complex system into parts which can be studied separately, and to identify combinations of variables which can be treated as a single variable. It is unfortunate that the problem statement cannot be generated by a large-scale digital computer.

English language statements are also used to clarify the desired operation of a logical computer; in particular, the operation of a relay control system is often written out in detail to aid in the understanding of the operation of the system. Expressing logical relations in sentences is in some cases very simple since the logical relations AND, OR, and NOT are closely related to the corresponding ordinary words. Ambiguity is possible in the case of OR, since there are two possible meanings of "A or B": first, the inclusive "or," which we have symbolized by OR, translated as "A or B or both"; second, the exclusive "or," which is read "A or B but not both." The other source of difficulty in sentences is that which is solved by the use of parentheses in equations. For example, the statement "X is on if B and C or D is on" could be interpreted either as $X=B\times C+D$ or $X=B\times(C+D)$. More complex statements cause increasing difficulty which can be overcome by defining intermediate variables or introducing parentheses or other punctuation in the sentences. The language techniques used by patent lawyers offer another possible approach which, however, will probably never become popular.

Two rather simple examples of logical expressions will now be given in a variety of representations. Words: "X is on if either B and C are both on or D is on"; "X' is on if either B and C are both on or D is on, but not both." A table of combinations follows:

B	C	D	X	X'
0	0	0	0	0
0	0	1	1	1
0	1	0	0	0
0	1	1	1	1
1	0	0	0	0
1	0	1	1	1
1	1	0	1	1
1	1	1	1	0

Karnaugh maps follow:

X

D \ B,C	00	01	11	10
0	0	0	1	0
1	1	1	1	1

X'

D \ B,C	00	01	11	10
0	0	0	1	0
1	1	1	0	1

Disjunctive normal form follows:

$$X=\bar{B}\bar{C}D+\bar{B}CD+B\bar{C}D+BC\bar{D}+BCD$$

$$X'=\bar{B}\bar{C}D+\bar{B}CD+B\bar{C}D+BC\bar{D}$$

Simplified expressions (readily obtained from the maps or from algebraic manipulation of the preceding expressions) follow:

$$X=B\times C+D$$

$$X'=(\bar{B}+\bar{C})\times D+B\times C\times\bar{D}$$

Finally, block diagrams are shown in Fig. 5. Block diagrams are of value because they can be used directly as schematic diagrams if the set of logical building blocks is sufficiently compatible.

None of these descriptions includes the effects of delays in the logical elements. However, after the equations and/or block diagrams have been derived, it is simple enough to analyze delays and introduce additional delays as needed for compatibility.

SIMPLIFICATION, ECONOMY, AND RELIABILITY

The problem of simplifying logical expressions has received a great deal of attention, and a number of techniques have been described in the literature.[11,12,13] They are based on a systematic application of certain identities in Boolean algebra. While their use is limited to perhaps seven or eight variables when the work is performed by human beings, the use of digital computers may be expected to extend this number and to reduce the time required.

Reducing the occurrences of variables may or may not lead to a reduction in cost. If one element (such as the OR) costs less than another (such as the AND), it may be preferable to increase the number of elements. Thus, the identity

$$AB+CD=(A+C)(A+D)(B+C)(B+D)$$

leads to the two block diagrams of Fig. 6. The form containing three logical elements may prove more expensive than that with five. If relative reliability figures can be assigned to the various elements, similar reasoning is applicable.

VERIFICATION

The most convincing evidence of the correct design of a system is a demonstration of its correct operation. Also, demonstration is perhaps the best method of showing the relay control system designer that his relay actually can be replaced by magnetic logic circuits with no moving parts. The apparatus used for such demonstrations is called a simulator, and consists of a large collection of logical building blocks which can be interconnected to form any switching circuit desired. Input variables can be controlled by toggle switches or push buttons, with pilot lights for indication. Similarly, output indication can be obtained with pilot lights or, for more dramatic demonstrations, the output signals can be used to control operating models of the actual apparatus to be controlled. Fig. 7 shows a typical simulator.

A most important use of the simulator applies to education. There is no simpler or quicker way than the use of the simulator of demonstrating the elementary logical operations, their simple combinations, and logical feedback circuits to personnel unfamiliar with switching circuits. Comparable simulators with relay elements are in use at Massachusetts Institute of Technology and Bell Telephone Laboratories for verification and education.

Verification of system design can also be performed mathematically if the compatibility properties of the logical elements have been well established. Some electronic digital computers which work in the binary number system have in their set of instructions the logical operations AND, OR, and NOT; and the memory can be used to simulate flip-flops. It is therefore possible to program such a computer to calculate the output for every set and sequence of inputs and to print out the results.

Verification of a large system can usually be accomplished by breaking it down into smaller parts. Often, design will have been started by making the same breakdown, and then verification can keep pace with design.

Conclusions

This paper discusses the logical properties of a set of magnetic logic circuits and presents a simple algebraic representation of their performance. The effects of the physical characteristics of magnetic logic circuits on various aspects of switching circuit design are noted. Finally, the advantages of simulators for verification and education are pointed out.

Magnetic logic circuits present the system designer with a number of new and interesting problems. In addition, there remain the standard unsolved problems of switching circuits: how to find the most economical circuit, and how to handle many variables. Ingenuity will continue to be a major tool in design work in the foreseeable future.

References

1. Magnetic Logic Circuits for Industrial Control Systems, **W. G. Evans, W. G. Hall, R. I. Van Nice.** *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 166–71.

2. The Design of Switching Circuits (book), **William Keister, A. E. Ritchie, S. H. Washburn.** D. Van Nostrand and Company, Inc., New York, N. Y., 1951.

3. A Symbolic Analysis of Relay and Switching Circuits, **Claude E. Shannon.** *AIEE Transactions*, vol. 57, 1938, pp. 713–23.

4. Element of Boolean Algebra for the Study of Information-Handling Systems, **R. Serrell.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, no. 10, 1953, pp. 1366–80.

5. Arithmetic Operations in Digital Computers (book), **R. K. Richards.** D. Van Nostrand and Company, Inc., New York, N. Y., 1955.

6. Pulse-Switching Circuits Using Magnetic Cores, **M. Karnaugh.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, no. 5, 1955, pp. 570–84.

7. Magnetic Elements in Arithmetic and Control Circuits, **I. L. Auerbach, S. B. Disson.** *Electrical Engineering*, vol. 74, Sept. 1955, pp. 766–70.

8. An Application of Boolean Algebra to the Design of Electronic Switching Circuits, **S. H. Washburn.** *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 380–88.

9. An Algebraic Theory for Use in Digital Computer Design, **E. C. Nelson.** *PGEC Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., vol. EC-3, no. 3, 1954.

10. The Single-Core Magnetic Amplifier as a Computer Element, **R. A. Ramey.** *AIEE Transactions*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 442–46.

11. Synthesis of Electronic Computing and Control Circuits (book), **Staff of the Harvard Computation Laboratory.** Harvard University Press, Cambridge, Mass., 1951.

12. The Problem of Simplifying Truth-Functions, **W. V. Quine.** *The American Mathematical Monthly*, Menasha, Wis., vol. 59, Oct. 1952.

13. A Chart Method for Simplifying Truth Functions, **E. W. Veitch.** *Proceedings*, Association for Computing Machinery, Pittsburgh, Pa., May 1952.

14. The Synthesis of Sequential Switching Circuits, **D. A. Huffman.** *Journal*, Franklin Institute, Philadelphia, Pa., vol. 257, no. 3, pp. 61–190, and no. 4, pp. 275–303, 1954.

15. The Map Method for Synthesis of Combinational Logic Circuits, **M. Karnaugh.** *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 593–99.

16. Minimizing and Mapping Sequential Circuits, **W. S. Bennett.** *Ibid.*, vol. 74, pt. I, Sept. 1955, pp. 443–47.

A Data Transmission Machine

C. R. DOTY
ASSOCIATE MEMBER AIEE

L. A. TATE
ASSOCIATE MEMBER AIEE

MODERN business operations have developed a requirement for rapid and accurate communications involving the exchange of engineering, scientific, statistical, and accounting data. The terminal points may be in the same building or separated by hundreds of miles. The Armed Forces have the same requirements but the distances may be greater, and radio transmission must be used where wire facilities are not available.

The requirements for a communication system for transmitting data are substantially different from the requirements for transmitting message traffic. The fundamental difference lies in the degree of accuracy required. In message traffic, the inherent redundancy of English text permits reading through an occasional error. By comparison, the inherent redundancy in data is extremely low, and a single error may cause misinterpretation of the message. For this reason, an error-detecting means must be used to obtain accurate data transmission.

A second difference lies in the means for recording the information. Message traffic usually is recorded as printed copy. Data usually are recorded for reading by other data-processing machines, which use either punched cards, magnetic tape, or paper tape.

Equipment designed for message traffic has been applied successfully to some forms of data transmission, by the use of paper tapes to operate automatic telegraph machines. Extensive communications systems of this type have been in operation since 1941, when card-to-tape and tape-to-card translating machines were marketed. However, extra steps of data processing usually are required in preparing and utilizing the paper tapes, and existing telegraph machines do not use error-detecting codes.

This paper describes a new machine, the Data Transceiver, which is designed to meet the requirements for rapid and accurate transmission of all types of data.

Paper 56-985, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted June 14, 1956; made available for printing August 7, 1956.

C. R. Doty and L. A. Tate are with International Business Machines Corporation, Poughkeepsie, N.Y.

Description

The IBM (International Business Machines) Data Transceiver reads and transmits information from punched cards, and records the information by punching cards at the receiving location. Transmission may be over telephone or telegraph circuits, using wire, radiotelegraph, or microwave.

As shown in Fig. 1, the machine is composed of two units. The unit on the right contains the card reading and punching mechanisms, the control keyboard, and circuits. The unit on the left contains the electronic commutator and circuitry for transmitting and receiving signals. Different types of signal units are used for telephone and telegraph operation because of the difference in signalling requirements of the two types of circuits.

Telegraph signal units may be operated at any of the standard keying speeds of 60, 75, or 100 words per minute, transmitting three, four, or five cards per minute if all 80 card columns are transmitted. The keying speed is determined by a pluggable speed selector, which may be changed by the operator to match the rated speed of the telegraph circuit.

Telephone signal units transmit and receive tone signals at the rate of 11 cards per minute if all 80 card columns are transmitted. Since this speed does not require the full bandwidth of a telephone circuit, the band is divided into four channels for the simultaneous operation of four pairs of machines. Thus, a single circuit can transmit 44 cards per minute. The channel separation is accomplished by band-pass filters, which are pluggable to permit the operator to change the channel assignment of a machine.

Error Detection and Correction System

The code used in punching cards contains 12 elements per card column, but only 50 alphabetical, numerical, and special characters are in common use. Four additional code combinations are used for signalling, but are not punched into the cards. The card code is translated to use fewer code elements, which permits more efficient use of the information capacity of the transmission line. A minimum of six code elements is required to make these 54 code combinations without using figure and letter shifts.

The error-detecting code is formed by adding two redundant code elements to each character so that each transmitted character contains four marking and four spacing code elements, as shown in Fig. 2. The receiving machine checks each character by counting the marking code elements. A transmission error is indicated if the received code combination is anything other than one of those listed in Fig. 2. Thus, the receiver can detect all single errors and all multiple errors, except those which compensate each other to produce no net change in the number of marking or spacing elements in a character. This error-detecting code was chosen on the basis of field-testing experience, to provide the required degree of accuracy with a minimum of redundancy added to the message. These tests were conducted on many types of communications circuits, including transoceanic radio.

In addition to the fixed-count and invalid-code combination checks of each character, the receiving machine performs other checking functions, including a check to see that the correct number of characters is received for each card, and that the machine is not out of cards. Detection of any failure stops both the transmitting and the receiving machines. The receiving machine stops immediately upon detecting a failure. The transmitter stops at the end of the card being transmitted when the failure occurred. So that the receiving machine can exercise complete control over the transmitting machine, the transmitter is arranged to stop at the end of each card and send a special code combination (end-of-card). If no failure has been detected the receiver responds by sending an acknowledging signal (start). Reception of this acknowledging signal permits the transmitter to feed the next card

Fig. 1. IBM Data Transceiver

CHARACTER	CARD CODE	TRANSMITTAL CODE							
SPACE	–			N	R	7	4		
A	12-1	X	O		R				1
B	12-2	X	O		R			2	
C	12-3	X	O					2	1
D	12-4	X	O		R		4		
E	12-5	X	O				4		1
F	12-6	X	O				4	2	
G	12-7	X	O		R	7			
H	12-8	X	O			7			1
I	12-9	X	O			7		2	
J	11-1	X		N	R				1
K	11-2	X		N	R			2	
L	11-3	X		N				2	1
M	11-4	X		N	R		4		
N	11-5	X		N			4		1
O	11-6	X		N			4	2	
P	11-7	X		N	R	7			
Q	11-8	X		N		7			1
R	11-9	X		N		7		2	
S	0-2		O	N	R			2	
T	0-3		O	N				2	1
U	0-4		O	N	R		4		
V	0-5		O	N			4		1
W	0-6		O	N			4	2	
X	0-7		O	N	R	7			
Y	0-8		O	N		7			1
Z	0-9		O	N		7		2	
&	12	X	O			7	4		
—	11	X		N		7	4		
0	0		O	N		7	4		
1	1	X	O	N					1
2	2	X	O	N				2	
3	3			N	R			2	1
4	4	X	O	N			4		
5	5			N	R		4		1
6	6			N	R		4	2	
7	7	X	O	N		7			
8	8			N	R	7			1
9	9			N	R	7		2	
$\overset{+}{0}$	12-0		O			7	4	2	
$\bar{0}$	11-0	X	O	N	R				
/	0-1		O	N	R				1
#	3-8				R	7		2	1
.	12-3-8			N		7		2	1
$	11-3-8	X				7		2	1
,	0-3-8		O			7		2	1
@	4-8				R	7	4		1
⌑	12-4-8			N		7	4		1
*	11-4-8	X				7	4		1
%	0-4-8		O			7	4		1
START OR END OF CARD		X				7	4	2	
RESTART						7	4	2	1
TELEPHONE TELEGRAPH					R	7	4	2	
END OF TRANSMISSION				N		7	4	2	

Fig. 2. IBM Data Transceiver, fixed count, four of eight code

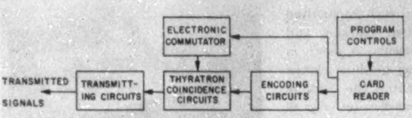


Fig. 3. Block diagram as a transmitter

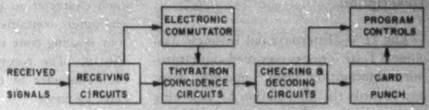


Fig. 4. Block diagram as a receiver

and resume transmission. If a failure is detected, and the receiver does not transmit the acknowledging signal, the transmitting machine stops. Both machines then must be restarted by the operators, and the transmission of the card is repeated. In this manner, detectable errors or other failures may be corrected immediately so that each received card is a complete and checked transmission.

SIGNALING CHARACTERISTICS

The 8-element code is transmitted in a 10.5-unit start-stop cycle. A synchronizing pulse, with a duration of one unit, is transmitted first, followed by the eight code elements and a stop interval which is at least 1.5 times the duration of a code element.

Polar relays are used for both transmission and reception of telegraph signals on either full-duplex or half-duplex neutral circuits. A machine may be operated at any of the standard keying speeds of 60, 75, or 100 words per minute (22.75, 28.4, or 37.1 dot-cycles per second). The speed is determined by a pluggable oscillator unit which may be changed by the operator.

Electronic circuitry permits the receiver to use extremely short samples of 50 microseconds for regenerating the received signals. Thus, distortion of the received signals will not interfere with the proper operation of the receiver, unless the distortion is so severe that it affects the center of a signal pulse. On circuits which use regeneration for printer operation, the regenerative repeaters must be disconnected during card transmission because of the different code characteristics.

For operation on telephone circuits, the transmitted signal is an audio-frequency carrier, the amplitude of which is modulated by keying the carrier on and off. When no signals are being transmitted the carrier is off, which permits the receiver to transmit acknowledging signals to the transmitter on the same carrier frequency. Signals are transmitted at the rate of 90 dot-cycles per second in a channel which occupies a bandwidth of 500 cycles per second. As many as four pairs of machines may be operated simultaneously on one telephone circuit, with each pair operating independently of the others on a carrier frequency of either 800, 1,300, 1,800, or 2,300 cycles per second. Band-pass filters are used for both transmission and reception to reduce interference from signals on the other channels.

Operation

TRANSMITTING

Fig. 3 is a block diagram of the Data Transceiver operating as a transmitter. Each card column which is to be transmitted, as determined by the program controls, is read by the electromechanical card reader. Electrical contacts in the reading mechanism operate relays through a diode encoding matrix. Contacts on the relays complete the translation from the card code to the transmittal code.

Operation of the card reader also starts the electronic commutator, which determines the timing of the transmitted signals. The timing reference is obtained from a clock oscillator, with different frequencies used for the different operating speeds. This oscillator starts and operates through ten cycles during the transmission of one character. At the end of this commutator cycle the oscillator is stopped. The first oscillator cycle is used to generate the synchronizing pulse. The next eight cycles determine the time sequence and duration of transmission of the eight code elements. The last oscillator cycle occurs during the stop interval.

The thyratron coincidence circuit compares the timing signals with the output of the encoder. Coincidence of a timing signal and the corresponding code output generates the serialized keying signal, which keys the transmitter. The transmitter is either a polar relay for operation on telegraph circuits, or a pulsed oscillator and band-pass filter for operation on telephone circuits.

RECEIVING

Fig. 4 is a block diagram of the Data Transceiver operating as a receiver. In telegraph operation, the received signals operate a polar relay. For telephone operation, the received signals pass through a band-pass filter, adjustable loss pad, amplifier, detector, and clipping amplifier. The line loss of the telephone circuit is compensated by the adjustment of the loss pad, while normal variations in the line loss are compensated by an automatic regulating circuit in the clipping amplifier.

The synchronizing signal, which precedes the eight code elements in each character, starts a cycle of the electronic commutator. The commutator cycle is not started, however, until the received signal is sustained for a 1/2-unit duration. This delay minimizes the probability that noise will cause false starting of the commutator. A receiving commutator has an operating cycle of ten units, but a transmitting commutator has a minimum cycle of 10.5 units. This time difference permits start-stop operation when the same clock oscillator frequency is used for timing both commutators.

The receiving commutator generates a series of nine sampling pulses, eight of which occur at the mid-point of the correct time of reception of the eight code elements. The coincidence of a received signal and a sampling pulse causes ionization of a thyratron, whose output current operates a coding relay.

The ninth sampling pulse occurs during the stop interval of the receiving commutator. This should occur in the stop interval between successive received characters. If a signal is received at this time, it may indicate that the receiving commutator is not in a proper synchronism with the received signals. Therefore, coincidence of a received signal and this ninth sampling pulse is treated as an error, and the receiving machine is stopped.

Contacts of the coding relays are used in the checking circuits for the detection of erroneous and invalid code combinations. Legitimate code combinations are translated into the card code by other contacts on the coding relays, and the output of the decoding circuit operates the card punching mechanism.

The same electronic commutator, thyratron coincidence circuits, and the coding relays are used for transmitting and receiving. Switching from one condition to the other is performed by relays which are operated by the program controls.

MISCELLANEOUS FUNCTIONS

Means are provided so that the operators may exchange co-ordinating signals

Vol. 7
Pt. 1
1956

to indicate that the receiving machine is ready for operation (start), that communication is desired on an alternate facility (telephone or telegraph), and that transmission is completed (end-of-transmission).

Other machine functions are used to adapt the Data Transceiver to established practices in using punched cards. The machine is controlled by programming the desired machine functions for each card column.

1. Card columns may be skipped, and skipping may be started in the same column of every card or by a control code punched in the card.
2. Information that is to be punched into all cards may be duplicated from one card to the next. Duplicated information is not transmitted.
3. During the time a receiving machine is skipping or duplicating, it cannot accept information from the transmitter. The transmitter may be stopped until the receiver is ready. At this time the receiver transmits a special code combination (restart) which causes the transmitter to resume transmission.
4. The machine normally stops on blank card columns to detect a failure of the card reader. This check may be disabled for any card columns in which blanks are permissible.
5. A machine may be equipped with a printing mechanism, which prints received or duplicated information on the card as the card is punched. Printing may be suppressed in selected card columns.

The operation of the machine is controlled by means of a program card, prepared by punching machine instructions in each column. This card is wrapped around a drum which rotates in step with the card-handling mechanism. Sensing contacts read the instructions and operate relays to control the machine. The transmitting and receiving programs need not be identical, but must be compatible, because the receiving machine will stop if an incorrect number of card columns is received, or if signals are received when the machine is not in the proper program condition.

Summary and Conclusions

The Data Transceiver may be operated on either wire or radio communication circuits. Punched cards are used for both input and output of data, and the machine has the programming flexibility required in using punched cards. The error detection and correction system supervises the operation of both machines. Field-testing and operating experience indicates that an extremely high degree of accuracy is obtained on all types of circuits.

The present uses of Data Transceivers and the increasing interest shown in them by commercial and military planners indicate that Data Transceivers will find a wide field of application in data transmission.

Synchronization Accuracy Obtainable with Multiplier Phototubes

LEO LEVI
ASSOCIATE MEMBER AIEE

Synopsis: The jitter in multiplier phototube synchronization systems is analyzed. The effect of bandwidth, phototube, and optical parameters on jitter is determined, and optimum operating parameters are derived. In the Appendix the intensity range of superiority of the multiplier phototubes over simple vacuum phototubes is determined.

ONE METHOD of obtaining accurate synchronization of rotating components with electronic signals is by the use of multiplier phototubes. Accuracies of the order of a millimicrosecond have been obtained with this method. The synchronization signal is obtained by scanning a light beam, usually a slit image, across a slit located in front of the phototube. The scanning is done by reflecting the light beam from a mirror fastened to the rotating component. A possible arrangement is shown in Fig. 1. [Light from the source is imaged on the source slit by means of the condensing lens L_1 (shown as two plano convex elements). Lens L_2 images the illuminated source slit on the phototube slit, if the rotating reflector is in the proper position. Lens L_3 images the exit pupil of L_2 on the photo cathode. As the reflector rotates, the slit image is scanned across the phototube slit and at the instant of coincident a light flash will strike the photo cathode. The resulting current pulse can be used for synchronization.]

As the object of the arrangement is synchronization, it is of great interest to know the amount of jitter introduced by the noise in the phototube. An attempt is made here to derive an expression for this jitter.

Since both the noise and the pulse rise-time are intimately connected with the bandwidth of the phototube circuit, the bandwidth required for minimum jitter is of interest also.

In the first part of this paper, arbitrarily fixed optical components are assumed and the bandwidth required to yield minimum jitter is determined. In the second part, optical components are taken into consideration as variables and their function in optimizing synchronization is determined.

Electronic Considerations

First an attempt will be made to derive the amount of jitter inherent in the circuit involving the phototube for any given optical rise time and light intensity.

Let the pulse rise time $=\tau$, the jitter $=\Delta t$, and the signal-to-noise ratio $=s$. Then it can be shown that

$$\Delta t=\tau/s \tag{1}$$

This may be derived from Fig. 2, where a pulse with rise time τ is illustrated. If the synchronization pulse is triggered at voltage V_t, it is apparent that the jitter $\Delta t=\overline{ec}$, when the noise $n=\overline{ed}$. Now, triangles *abc* and *ced* are similar and the following can be written

$$s=\frac{V_t}{n}=\frac{\overline{cb}}{\overline{ed}}=\frac{\overline{ab}}{\overline{ec}}=\frac{\tau}{\Delta t}$$

from which equation 1 follows.

Assume, letting Δf equal the bandwidth,

$$s=K/\sqrt{\Delta f} \tag{2}$$

Paper 56-759, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted October 3, 1955; made available for printing May 2, 1956.

Leo Levi is with the Fairchild Controls Corporation, New York, N. Y.

The results reported here were obtained in part under contract with the Evans Signal Corps Laboratory at Belmar, N. J.

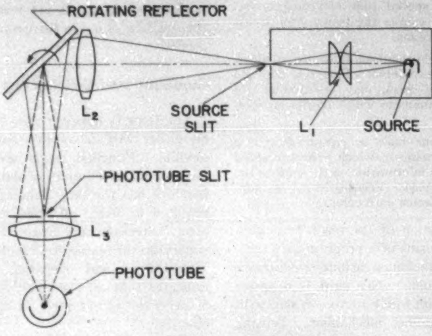


Fig. 1. Synchronization system-optical schematic

where K is signal-to-noise at one cycle-per-second bandwidth. This is based on the assumption of white noise where noise energy is uniformly distributed over the frequency band and therefore noise voltage squared is proportional to bandwidth. This assumption has been verified experimentally and is true for frequencies up to 10^8cycle-per-second.[1] Since $s \propto 1/\text{noise}$, equation 2 follows. Nothing further is assumed concerning noise at this point.

A second assumption will be that the electrical rise time is inversely proportional to bandwidth, i.e.

$$\tau_e = \frac{k}{\Delta f} \tag{3}$$

where k is usually taken to be 0.35.[2]

Assume further that the rise times will combine in the following manner

$$\tau = \sqrt{\tau_0^2 + \tau_e^2} \tag{4}$$

where τ_0 is the optical rise time. This is based on Laplace's "central limit theorem" of probability.[3] Substituting equations 2, 3, and 4 into 1 gives

$$\Delta t = \frac{\sqrt{\Delta f}}{K}\sqrt{\tau_0^2 + \frac{k^2}{\Delta f^2}} = \frac{1}{K}\left(\tau_0^2 \Delta f + \frac{k^2}{\Delta f}\right)^{1/2} \tag{5}$$

Differentiating and equating to zero gives

$$\frac{d\Delta t}{d\Delta f} = \frac{1}{2K}\left(\tau_0^2 \Delta f + \frac{k^2}{\Delta f}\right)^{-1/2}\left(\tau_0^2 - \frac{k^2}{\Delta f^2}\right)$$

$$= 0 = \tau_0^2 - \frac{k^2}{\Delta f^2}$$

and the bandwidth for minimum jitter

$$\Delta f = \frac{k}{\tau_3} = \frac{0.35}{\tau_0} = \frac{0.35}{\tau_e} \tag{6}$$

(the last member from equation 3)

It is seen therefore that $\tau_0 = \tau_e$ and $\tau = \sqrt{2\tau_0}$, and it may be concluded that the minimum jitter resulting will be

$$\Delta t_{\min} = \left(\sqrt{\frac{0.35}{t_0}}/K\right)\sqrt{2}\tau_0$$

$$= \frac{\sqrt{0.7\tau_0}}{K} = 0.837\frac{\sqrt{\tau_0}}{K} \tag{7}$$

Up to this point nothing has been assumed concerning the noise except that it be "white" and on this basis the following conclusions have been arrived at:

1. For minimum jitter the optimum bandwidth is that which makes the electronic rise time equal to the optical rise time.
2. The jitter is then directly proportional to the square root of the optical rise time and inversely proportional to the signal-to-noise ratio at unity bandwidth.

The value of the noise figure K will now be determined. In general, noise in phototubes consists of shot noise due to the "graininess" of the electronic emission, Johnson noise in the load resistor, thermionic emission, and the ohmic leakage current.

In high bandwidth units, light levels are usually low (cf. the narrow slit necessitated by low optical rise time) and for this reason multiplier phototubes are imperative.[4,5] Since at room temperature the Johnson noise is negligible when the voltage drop across the load resistor is much more than 50 millivolts and, as a rule, this condition is met in the case of the multiplier phototubes, Johnson noise will be neglected here.

Similarly, it is assumed that operation will be at light levels considerably in excess of the "equivalent noise input" so that thermionic and leakage currents, too, may be neglected. This leaves only shot noise to be considered.

Shot noise from a single cathode is given by

$$i^2 = 2eI_0\Delta f \tag{8}$$

where i is the noise current, I_0 is the average cathode current, e is the charge of the electron and Δf is the bandwidth of the associated circuitry. This expression holds for $1/\Delta f$ large compared to the electron transition time.

When multiplier phototubes are used, the shot noise from the photo cathode is multiplied along with the original signal, so that the signal-to-noise ratio remains essentially constant. It is somewhat decreased by shot noise in successive dynodes. This reduction factor can be shown to be equal to

$$\sqrt{\frac{m-1}{m-1/M}}$$

where m is the multiplication at one dynode and $M = m^n$ is the total multiplication factor.[5,6]

Another factor decreasing the signal-to-noise ratio is imperfect electron focusing. This produces nonuniform gain over the

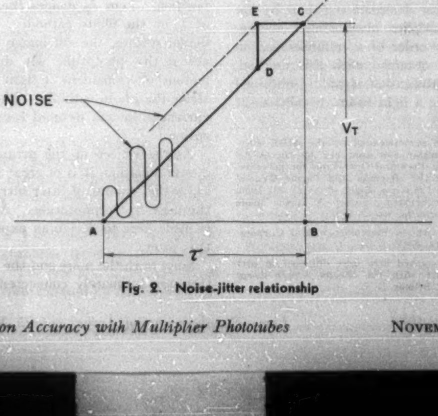


Fig. 2. Noise-jitter relationship

cathode area and consequently multiplies the noise by the rms of M which is more than the mean M. Thus the signal-to-noise ratio is degraded further by the factor[7]

$$\sqrt{\frac{\overline{M}^2}{\overline{\overline{M^2}}}}$$

Combining these last two factors, letting

$$K'^2 = \frac{m-1}{m-1/\overline{M}} \times \frac{\overline{M}^2}{\overline{\overline{M^2}}} \qquad (9)$$

an expression for the square of the signal-to-noise ratio of a multiplier photo tube is obtained from equations 8 and 9:

$$s^2 = \frac{I_0^2}{i^2} = \frac{I_0}{2e\Delta f} K'^2 = \frac{K^2}{\Delta f} \qquad (10)$$

the last member coming from equation 2.

It is concluded that

$$K = K'\sqrt{\frac{I_0}{2e}} \qquad (11)$$

But $I_0 = \sigma L$ where σ is the cathode luminous sensitivity (amperes per lumen) and L is the luminous flux, and therefore

$$K = K'\sqrt{\frac{\sigma L}{2e}} \qquad (12)$$

Optical Considerations

Now, consider the optical parameters, i.e., the velocity of the slit image v, the trigger slit width d, its height h, the distance from the rotating mirror to the trigger slit D, and the area of the exit pupil A of the imaging lens.

The total flux of the photocell[8] is given by

$$L = \pi B \sin^2 \theta \eta A_s = B\frac{A}{D^2}\eta h d \qquad (13)$$

where B is the source brightness (candles per square millimeter); A_s is the slit area $h \times d$; θ is the angle subtended by the radius of the exit pupil, area A, at the slit; and η is the optical efficiency of the system (taking account of stray reflection and absorption of light). The third member is obtained by substituting

$$\sin^2 \theta = \frac{r^2}{D^2} = \frac{A}{\pi D^2}$$

Therefore

$$K^2 = K'^2 \frac{\sigma B A \eta h d}{2eD^2} = C_1 \frac{d}{D^2} \qquad (14)$$

where

$$C_1 = \frac{1}{2}\sigma B A \eta h K'^2/e \qquad (15)$$

Also, from the geometry of Fig. 1 and assuming a perfect slit image

$$\tau_0 = \frac{d}{v} = d/2\pi\omega D = C_2\frac{d}{D} \qquad (16)$$

where $C_2 = 1/2\pi\omega$ and ω = angular velocity of the beam = twice the angular velocity of the reflector.

Then

$$\Delta t_{\min}^2 = 0.7\tau_0/K^2 = \frac{0.7C_2 d/D}{C_1 d/D^2} = C_3^2 D$$

and it is concluded that

$$\Delta t_{\min} = C_3\sqrt{D} \qquad (17)$$

where

$$C_3 = \sqrt{\frac{0.7C_2}{C_1}} = \sqrt{\frac{0.7 \times 2e}{2\pi\omega\sigma B A \eta h K'^2}} = \frac{\sqrt{0.7}}{\sqrt{\pi}K'} \times \frac{\sqrt{e}}{\sqrt{\omega\sigma B A \eta h}} = \frac{5 \times 10^{-10}}{\sqrt{\omega\sigma B A h}} \qquad (18)$$

taking $e = 1.6 \times 10^{-19}$ and usual values for K' and η (approximately 0.5 for both). This result, equation 17, namely that the jitter is directly proportional to the square root of the optical path length, is significant. The engineer might be tempted to increase the path length for increased velocity of the slit image to obtain more accurate triggering, but this gain is in fact vitiated by the reduced light intensity.

Also interesting is the fact that the jitter is independent of the slit width, assuming, of course, that this is wide enough to put the phototube into the proper operating range. It should be noted that this derivation assumes a perfect optical design, and a light source of any desired dimension.

Conclusions

ELECTRONIC

It was found that the bandwidth of the circuitry associated with the multiplier phototube should be such that the rise time of a square pulse fed into it should equal the optical rise time, i.e., bandwidth = 0.35/optical rise time.

Under these conditions the resulting trigger jitter will be directly proportional to the square root of the optical rise time and inversely proportional to the signal-to-noise ratio at unity bandwidth. This will be true for regions where the noise may be considered white.

OPTICAL

With a given optical system, light source, multiplier photo tube, and angular velocity of reflector, the jitter will be proportional to the square root of the distance from the reflector to the phototube slit. This will be true if operation is at a light level sufficiently high so that only shot noise is important, and sufficiently low to maintain the tube in its linear operating range which extends over a region 1,000,000:1.[4]

The slit width used appears, then, to have no effect on jitter. It is limited only by other design parameters and its upper limit is determined by space limitations, available source area, and relative lens apertures. The lower limit is set by the circles of confusion which make up the slit image. (As a result of lens aberrations and diffraction effects, a lens images an object point as a disk. This disk is called circle of confusion.) When the diameter of these circles becomes comparable in size to the slit image, the quantitative optical analysis breaks down, because the optical rise time is then no longer given by equation 16 nor the image brightness by equation 13.

Also, it is seen that increasing any of the following factors will decrease jitter by a factor proportional to the square root of the increase.

1. Source brightness.
2. Angular velocity of the reflector.
3. Area of the exit pupil.
4. Height of slit (or slit image).
5. Photo cathode sensitivity.

Since other factors affecting rise time and jitter have been ignored, it is important that the associated electronic circuitry be reasonably noisefree and that a good lens design providing a well-corrected slit image be used. Also the near stationary exit pupil should be imaged on the photo cathode, so that a comparably stationary area there is illuminated. In this way, effects of nonuniformities of the cathode sensitivity are minimized.

Appendix

The illumination range over which a multiplier phototube is superior to a simple phototube may be determined as follows:

It is seen that the use of a multiplier phototube reduces considerably the effect of Johnson noise, but aggravates the shot noise somewhat.

The break-even point, i.e., the light level at which the signal-to-noise obtainable with the two types of tube is the same, may readily be determined as follows. The signal-to-noise ratios obtainable from simple and multiplier phototubes, respectively, are

$$s_0 = \frac{RI}{n_j + n_s} = \left(\frac{MRI}{Mn_j + Mn_s}\right) \qquad (19)$$

and

$$s_m = \frac{MRI}{n_j + \frac{Mn_s}{K'}} \qquad (20)$$

where R is the output resistance of the phototube, n_j is the Johnson noise, independent of cathode current I, and n_s is the shot noise proportional to $\sqrt{I}$. $1/K'$ is, as before, a factor representing the increase in noise in a multiplier phototube due to various minor factors such as shot noise at the dynodes and nonuniform multiplication across the cathodes' surface.

Clearly, the break-even point occurs when

$$Mn_j + Mn_s = n_j + \frac{1}{K'}Mn_s$$

or

$$\frac{n_j}{n_s} = \frac{\frac{1}{K'} - 1}{1 - \frac{1}{M}} \doteq \frac{1}{K'} - 1 \qquad (21)$$

If it is again assumed that $K' = 0.5$, the break-even point occurs when $n_j = n_s$. But $n_j^2 = 4kTR\Delta f$ and $n_s^2 = 2R^2eI\Delta f$, so that the foregoing requirement reduces to

$$\frac{2kT}{eR} = I = \sigma L$$

or

$$L = \frac{2kT}{\sigma eR} = 5.14 \times 10^6/R\sigma \text{ lumen} \qquad (22)$$

at room temperature, where $T = 300$ (Boltzmann's constant $k = 1.37 \times 10^{-16}$ and $e = 1.6 \times 10^{-19}$).

The maximum value of R will, of course, be determined by the interelectrode capacitances of the phototube and the bandwidth requirement.

If the design is limited to $R = \frac{1}{2\pi fC}$, equation 22 reduces to

$$L = 3.23 \times 10^6 fC/\sigma \qquad (23)$$

This result shows that in all practical cases a multiplier phototube will provide the higher signal-to-noise ratio as long as its anode ratings permit it to be operated with appreciable multiplication.

References

1. Calculated Frequency Spectrum of the Shot Noise from a Photo-Multiplier Tube, **R. D. Sard.** *Journal of Applied Physics*, New York, N. Y., vol. 17, 1946, pp. 768–77.

2. Vacuum Tube Amplifiers. *Radiation Laboratory Series*, Massachusetts Institute of Technology, Cambridge, Mass., McGraw-Hill Book Company, Inc., New York, N. Y., vol. 18, 1948, Rule 4, p. 80.

3. *Ibid.*, Rule 1, p. 77.

4. Multiplier Phototube Characteristics: Application to Low Light Levels, **R. W. Engstrom.** *Journal of the Optical Society of America*, New York, N. Y., vol. 37, 1947, pp. 420–31 (extensive bibliography).

5. The Secondary Emission Multiplier, **V. K. Zworykin, G. A. Morton, L. Malter.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 24, 1936, pp. 351–75.

6. A Theory of Noise for Electron Multipliers, **W. Shockley, J. R. Pierce.** *Ibid.*, vol. 26, 1938, pp. 321–32.

7. Photoelectric Multipliers, **S. Rodda.** *Journal of Scientific Instruments and of Physics in Industry*, London, England, vol. 26, 1949, pp. 65–70.

8. The Principles of Optics, **Hardy and Perrin.** McGraw-Hill Book Company, Inc., New York, N. Y., 1932, equation 223, p. 411.

Impedance of Thin-Wire Loop Antennas

JAMES E. STORER
NONMEMBER AIEE

THE THIN-wire loop is one of the first antennas to receive theoretical consideration, having been discussed by Pocklington[1] in 1897. Pocklington treated a closed loop excited by a plane wave; he obtained an exact solution for the current on the loop in the form of a Fourier series. More recently, Hallén[2] considered a driven loop and obtained a solution, again in the form of a Fourier series, for the current and the impedance. However, Hallén pointed out that the coefficients of this series contained a singularity which made the series only quasiconvergent and hence useful only for loops small in comparison to a wave length. Moreover, the individual terms were complicated and their evaluation and a summation involved a somewhat difficult numerical task.

More recently, in an effort to obtain numerical results, other authors have dealt with the problem using approximation methods. Chang,[3] for example, applied the Hallén-King-Middleton expansion, Schelkunoff[4] has used a guided mode approximation; and the author has used a variational approach (unpublished). All of these approximation methods have one feature in common; they yield results which are in good agreement qualitatively with experiment, but poor agreement quantitatively. (It is quite possible that all of these methods, particularly Chang's could be made to yield better results by going to higher degrees of approximation; the resulting numerical labor, however, is likely to be prohibitive.) The reason for this can be explained by noting that all the approximation methods require some assumption as to the current distribution around the loop. The most common assumption is that the current distribution approximates a sinusoidal distribution. As will be shown subsequently, the sinusoidal assumption is not satisfactory, particularly for the current near the driving point of the antenna.

In this paper the rigorous Fourier series solution obtained by Hallén is reexamined, and modified so that the convergence difficulties which he encountered are avoided. Extensive numerical results are presented in Appendix I for the impedances of loops for varying wire sizes and circumferences up to $2^1/_2$ wave lengths. Appendix II presents some curves which aid in the computation of field patterns and current distributions. For an antenna having a particular wire size, some experimentally measured impedances are presented which agree well with theory.

Fourier Series Solution for the Current Distribution

Integral equations for the current distribution on thin-wire antenna structures are readily obtained by expressing the electric field as a function of the current, through Helmholtz integrals, and then equating the total electric field to zero along the wire surface. Following this procedure, with harmonic time dependence of the form $\epsilon^{+j\omega t}$, and with co-ordinate system and dimensions as indicated in Fig. 1, the integral equation for the circular loop antenna can be written as

$$V\delta(\phi) = \frac{j\zeta_0}{4\pi}\int_{-\pi}^{\pi} K(\phi-\phi')I(\phi')d\phi' \qquad (1)$$

where $I(\phi)$ is the total current at ϕ on the loop; V is the voltage of the slice generator exciting the loop at $\phi=0$; $\delta(\phi)$ is the

---

Paper 56-692, recommended by the AIEE Special Communications Applications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted October 28, 1955; made available for printing June 6, 1956.

Dr. James E. Storer is with Harvard University, Cambridge, Mass.

The author wishes to thank Prof. R. W. P. King of Harvard University for his help and encouragement with this research; Miss Phyllis Kennedy for making her experimental measurements available; and to Leon Levy, who performed the numerical computations. The author also wishes to acknowledge the financial support he received from the Atomic Energy Commission as a Fellow during the early phase of this research; and during the latter part, under the sponsorship of the Office of Naval Research, the Signal Corps of the U. S. Army, and the U. S. Air Force *Contract N5ori-76*. This research was originally published as *Technical Report No. 212*, Cruft Laboratory, Harvard University, May 1, 1955.

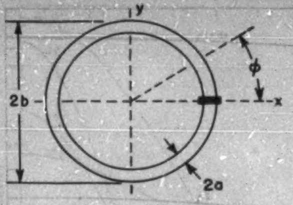


Fig. 1. Co-ordinates for loop antenna

Dirac delta function; and $k=\omega/c=2\pi/\lambda$; $\zeta_0=\sqrt{\frac{\mu_0}{\epsilon_0}}=120\pi$ ohms. The kernel of the integral equation 1 is given explicitly by

$$K(\phi-\phi')=\left[kb\cos(\phi-\phi')+\frac{1}{kb}\frac{\partial^2}{\partial\phi^2}\right]\times\frac{\epsilon^{-jkbR(\phi-\phi')}}{R(\phi-\phi')} \qquad (2A)$$

$$R(\phi-\phi')=\left[4\sin^2\left(\frac{\phi-\phi'}{2}\right)+a^2/b^2\right]^{1/2} \qquad (2B)$$

where a is the radius of the wire and b is the radius of the loop.

The thin-wire assumption, which provides the basis for obtaining this one-dimensional current equation, can be expressed explicitly as $a^2<<b^2$, $k^2a^2<<1$. The resulting solution cannot be more accurate than the order of these approximations.

Since $\frac{1}{R(\phi-\phi')}\epsilon^{-jkbR(\phi-\phi')}$ is a bounded, periodic function of ϕ, it can be expanded into a Fourier series, i.e.

$$\frac{1}{R(\phi-\phi')}\epsilon^{-jkbR(\phi-\phi')}=\sum_{-\infty}^{\infty}K_n\epsilon^{-jn(\phi-\phi')} \qquad (3)$$

$$K_n=K_{-n}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\epsilon^{-jkbR(\phi)}}{R(\phi)}\epsilon^{-jn\phi}d\phi \qquad (4)$$

Using equation 3 together with equation 2, it is seen that

$$K(\phi-\phi')=\sum_{-\infty}^{\infty}\alpha_n\epsilon^{jn(\phi-\phi')} \qquad (5)$$

where

$$\alpha_n=\alpha_{-n}=kb\left\{\frac{K_{n+1}+K_{n-1}}{2}\right\}-\frac{n^2}{kb}K_n \qquad (6)$$

Inserting expression 5 into integral equation 1 yields

$$V\delta(\phi)=\frac{j\zeta_0}{4\pi}\sum_{-\infty}^{\infty}\alpha_n\int_{-\pi}^{\pi}\epsilon^{jn(\phi-\phi')}I(\phi')d\phi' \qquad (7)$$

After expanding $I(\phi)$ into a Fourier series

$$I(\phi)=\sum_{-\infty}^{\infty}I_n\epsilon^{jn\phi};\quad I_n=\frac{1}{2\pi}\int_{-\pi}^{\pi}I(\phi)\epsilon^{-jn\phi}d\phi \qquad (8)$$

it can be seen that equation 7 reduces to the following

$$V\delta(\phi)=\frac{j\zeta_0}{2}\sum_{-\infty}^{\infty}\alpha_nI_n\epsilon^{jn\phi}$$

Hence

$$I_n=\frac{1}{j\pi\zeta_0\alpha_n}\int_{-\pi}^{\pi}\epsilon^{-jn\phi}V\delta(\phi)d\phi=\frac{V}{j\pi\zeta_0\alpha_n}$$

Inserting this result into equation 8 yields

$$I(\phi)=\frac{V}{j\pi\zeta_0}\sum_{-\infty}^{\infty}\frac{\epsilon^{jn\phi}}{\alpha_n}=\frac{V}{j\pi\zeta_0}\times\left(\frac{1}{\alpha_0}+2\sum_{1}^{\infty}\frac{\cos n\phi}{\alpha_n}\right) \qquad (9)$$

From this, the impedance of the antenna Z is found to be

$$Z=\frac{V}{I(0)}=j\pi\zeta_0\left(\frac{1}{\alpha_0}+2\sum_{1}^{\infty}\frac{1}{\alpha_n}\right)^{-1} \qquad (10)$$

These results, equations 9 and 10, obtained by Hallén, constitute a formal solution of the loop antenna. From them the transmitting pattern and by reciprocity the receiving pattern can be obtained. However, in order to make them useful, some way must be found to evaluate the series numerically.

It can be shown that these equations, 9 and 10, are in agreement with the theory of small loops. If equation 4 is used and the explicit evaluation of K_n given in Appendix III, it is readily shown that, for loops small in comparison to the wave length, the current is nearly a constant, independent of ϕ and

$$kb\ll 1,\ Z\cong j\pi\zeta_0\alpha_0=j\pi\zeta_0kbK_1$$

$$\cong\frac{\pi\zeta_0}{6}k^4b^4+j\pi\zeta_0kb\left(\ln\frac{8b}{a}-2\right)$$

This is the usual equation for the resistance and reactance of a small loop.

The Fourier Series

It is apparent from the preceding derivation that the usefulness of this method of solution depends on the evaluation of the series

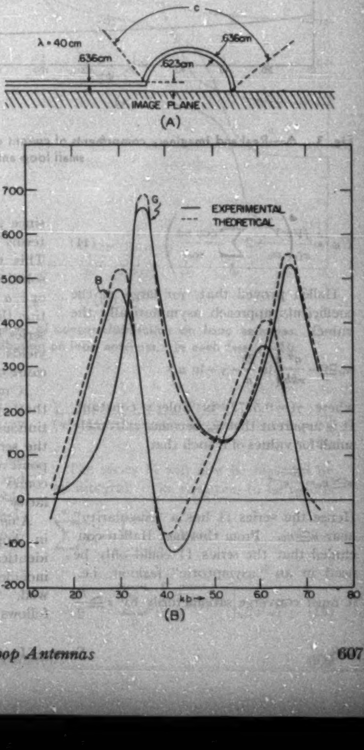


Fig. 2. A—Experimental dimensions. B—Impedance of loop antenna

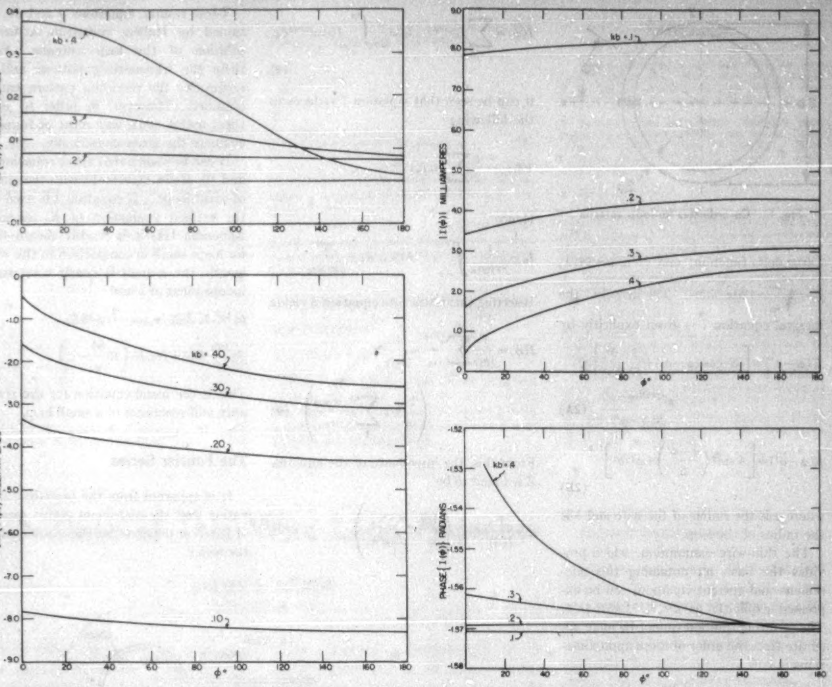


Fig. 3. A—Real and imaginary components of current distribution on small loop antennas; B—Magnitude and phase of current distribution on small loop antennas. In each case $\Omega = 2\ln(2\pi b/a) = 10$

$$I(\phi)=\frac{jV}{\pi\zeta_0}\left(\frac{1}{\alpha_0}+2\sum_{1}^{\infty}\frac{\cos n\phi}{\alpha_n}\right) \qquad (11)$$

Hallén proved that, for large n, the coefficients approach asymptotically the value.

$$\alpha_n \cong -\frac{n^2}{\pi kb}\left\{\ln\frac{2b}{a}-\gamma-\ln n\right\}$$

where $\gamma(=0.5772)$ is Euler's constant. It is apparent that α_n becomes extremely small for values of n such that

$$n \cong n_0=\frac{2b}{a}\epsilon^{-\gamma}$$

Hence the series 11 has a "singularity" near $n \cong n_0$. From this fact Hallén concluded that the series 11 could only be used in an "asymptotic" fashion, i.e., it must converge satisfactorily by $r \cong \frac{n_0}{2}$ since' after this, the value of the individual terms begin increasing in magnitude. This restriction meant that the series solution 11 was only useful for kb small or b/a very large. Even with this limitation the summation of $n_0/2$ terms of this series in a formidable task and, at best, yields relatively inaccurate results because of the "singularity."

It must be remembered at this point that current is both bounded and continuous (for physical reasons) and hence the series 11 must converge. When this point of view is adopted, the problem becomes one of treating Hallén's "singularity" in a more rigorous fashion.

A derivation of the value of αn is given in Appendix III, which is essentially identical to that of Hallén's, but which includes the dominant complex term as well. The result for large n is given as follows

$$\alpha_n \sim \frac{1}{\pi}\left(kb-\frac{n^2}{kb}\right)\left(\ln n_0-\ln n-j\times\frac{(kb)^{2n+1}}{\Gamma(2n+2)}\right) \quad \begin{matrix} n>kb \\ n\gg 1\end{matrix} \qquad (12)$$

where $n_0=\frac{2b}{a}\epsilon^{-\gamma}$

It is apparent that the inclusion of the rather negligible complex term in equation 12 cannot alter significantly the sum of the resulting series. However, with its inclusion α_n is never equal to zero. This fact will be used subsequently to permit a replacement of the series by an integral.

The following work will be restricted to loops in which $kb<2.5$, i.e., the circumference of the loop is less than $2\frac{1}{2}$ wave lengths. (The derivation can readily be modified to include values of kb larger than 2.5 if desired.) Almost all loop antennas of practical interest are contained

Vol. 7
Pt. 1
1956

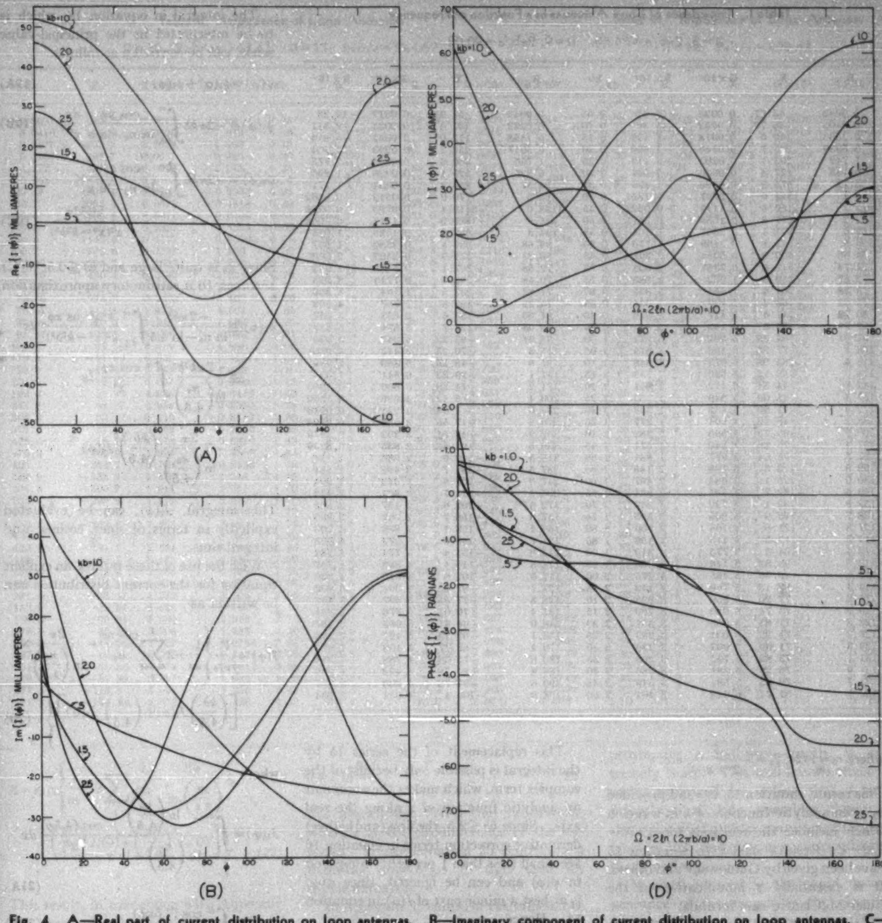


Fig. 4. A—Real part of current distribution on loop antennas. B—Imaginary component of current distribution on loop antennas. C—Magnitude of current distribution on loop antennas. D—Phase of current distribution on loop antennas. In each case $\Omega=10$

in this range. The series 11 can then be written in the form

$$I(\phi)=\frac{V}{j\pi\zeta_0}\left(\frac{1}{\alpha_0}+2\sum_{1}^{4}\frac{\cos n\phi}{\alpha_n}+\psi(\phi)\right) \qquad (13)$$

where

$$\psi(\phi)=2\sum_{5}^{\infty}\frac{\cos n\phi}{\alpha_n}$$

The procedure to be used will sum the first five terms of the series explicitly, and replace the remainder of series $\psi(\phi)$ by an integral.

Now, it can be shown by an insertion of numerical values that for $kb \leqq 2.5$ and $n \geqq 5$, the value of α_n differs negligibly from the asymptotic value given by equation 12. Hence, to an excellent approximation

$$\psi(\theta)=-2\pi kb\sum_{5}^{\infty}\times$$

$$\frac{\cos n\phi}{(n^2-k^2b^2)[\ln n_0-\ln n-j(kb)^{2n+1}/\Gamma(2n+2)]} \qquad (14)$$

The series 14 will now be replaced by an integral. The equation to be used is

$$\sum_{N}^{\infty}a_n=\int_{N-1/2}^{\infty}a_x dx+\sum_{0}^{\infty}c_n\left[\frac{d^{2n+1}}{dx^{2n+1}}a_x\right]_{x=N-1/2} \qquad (15)$$

Table I. Impedance of Loop Antennas as a Function of Frequency

$\Omega=8$, $2\pi b/a=54.60$. $\Omega=9$, $2\pi b/a=90.02$

R	X	$G\times10^3$	$B\times10^3$	kb	R	X	$G\times10^3$	$B\times10^3$
0.0048	43.57	0.0025	−22.95	0.05	0.0046	51.99	0.0017	−19.23
0.0402	88.77	0.0052	−11.36	0.10	0.0392	107.4	0.0034	−9.311
0.1533	140.5	0.0078	−7.119	0.15	0.1538	172.0	0.0052	−5.814
0.5939	205.7	0.0140	−4.860	0.20	0.5917	252.2	0.0093	−3.964
1.742	293.1	0.0203	−3.412	0.25	1.756	360.6	0.0135	−2.773
6.143	427.4	0.0336	−2.339	0.30	6.327	529.1	0.0226	−1.890
23.72	675.2	0.0520	−1.479	0.35	25.57	853.2	0.0351	−1.171
140.1	1344.0	0.0767	−0.7361	0.40	162.5	1776.0	0.0518	−0.5662
7972.7	2189.8	0.1172	−0.0322	0.45	1188.0	−3119.4	0.0796	0.0209
502.2	−1677.8	0.1638	0.5471	0.50	415.0	−1887.9	0.1111	0.5054
169.0	−824.4	0.2387	1.164	0.55	154.3	−962.0	0.1625	1.013
106.6	−544.1	0.3469	1.770	0.60	99.68	−639.4	0.2380	1.527
84.56	−400.2	0.5054	2.392	0.65	80.23	−471.9	0.3502	2.060
73.74	−311.8	0.7357	3.028	0.70	72.58	−367.1	0.5183	2.622
72.87	−250.3	1.072	3.683	0.75	70.48	−293.3	0.7744	3.223
73.28	−204.4	1.554	4.335	0.80	71.62	−237.5	1.164	3.860
76.02	−168.4	2.227	4.934	0.85	75.22	−192.8	1.757	4.502
80.72	−141.4	3.117	5.376	0.90	81.00	−155.6	2.633	5.057
87.59	−115.5	4.193	5.531	0.95	89.08	−122.7	3.874	5.337
94.81	−94.30	5.302	5.274	1.00	98.94	−95.22	5.247	5.050
104.4	−77.20	6.193	4.579	1.05	111.8	−70.25	6.413	4.030
115.4	−63.08	6.672	3.648	1.10	127.6	−48.16	6.860	2.589
128.0	−52.23	6.696	2.732	1.15	147.1	−29.22	6.541	1.300
142.0	−44.54	6.412	2.011	1.20	171.2	−14.36	5.800	0.4866
157.4	−41.09	5.949	1.553	1.25	200.4	−5.427	4.986	0.1350
172.9	−41.89	5.462	1.323	1.30	234.0	−4.986	4.271	0.0910
188.1	−48.12	4.991	1.277	1.35	270.2	−17.55	3.686	0.2394
200.7	−60.73	4.565	1.382	1.40	302.2	−46.51	3.232	0.4975
207.4	−76.86	4.240	1.571	1.45	320.5	−91.53	2.885	0.8239
207.5	−95.54	3.977	1.831	1.50	315.5	−142.0	2.636	1.186
199.8	−113.3	3.788	2.148	1.55	287.5	−184.7	2.462	1.581
186.0	−126.1	3.684	2.497	1.60	247.0	−207.6	2.373	1.994
169.5	−132.7	3.657	2.864	1.65	206.3	−211.7	2.361	2.423
152.8	−133.1	3.722	3.242	1.70	172.5	−202.6	2.437	2.862
138.7	−128.9	3.869	3.595	1.75	147.8	−186.7	2.607	3.293
127.8	−121.6	4.106	3.909	1.80	130.9	−168.1	2.883	3.703
119.5	−112.6	4.415	4.158	1.85	121.7	−149.6	3.273	4.023
115.4	−104.2	4.775	4.310	1.90	114.6	−131.4	3.774	4.282
113.2	−95.82	5.146	4.356	1.95	114.8	−115.1	4.345	4.357
113.4	−88.30	5.485	4.275	2.00	117.0	−101.6	4.915	4.226
114.4	−82.18	5.766	4.142	2.05	121.7	−87.95	5.398	3.902
116.8	−77.20	5.958	3.938	2.10	128.8	−77.59	5.697	3.432
119.9	−73.74	6.050	3.720	2.15	137.4	−70.01	5.776	2.941
123.2	−71.84	5.056	3.531	2.20	148.0	−64.94	5.665	2.485
126.2	−71.22	6.001	3.385	2.25	159.1	−63.49	5.421	2.163
129.3	−71.99	5.903	3.286	2.30	170.4	−65.77	5.108	1.972
131.1	−73.82	5.792	3.263	2.35	180.0	−72.08	4.788	1.918
131.9	−76.29	5.681	3.286	2.40	186.8	−81.78	4.492	1.966
131.7	−79.05	5.581	3.349	2.45	190.6	−94.78	4.206	2.091
130.1	−81.71	5.512	3.461	2.50	178.8	−106.4	4.031	2.284

where $c_0=1/24$, $c_1=-\dfrac{7}{2^4\times360}$, etc.

This result, equation 15, is valid provided a_n is an analytic function of n in a region which includes the real axis for $n>N-1/2-\epsilon$. Results similar to equation 15 have been given by Gumowski[5] and others. It is essentially a modification of the Euler-McClaurin sum formula.

Using equation 15 in connection with equation 14 yields

$$\psi(\phi)=-2\pi kb\int_{4.5}^{\infty}\times\frac{\cos x\phi dx}{(x^2-k^2b^2)\left(\ln n_0-\ln x-j\dfrac{(kb)^{2n+1}}{\Gamma(2x+2)}\right)}\tag{16}$$

$$-\frac{2\pi kb}{24}\times\left[\frac{d}{dx}\frac{\cos x\phi}{(x^2-k^2b^2)\left(\ln n_0-\ln x-j\dfrac{(kb)^{2x+1}}{\Gamma(2x+2)}\right)}\right]_{x=4.5}-\cdots$$

This replacement of the series 15 by the integral is possible only because of the complex term, which makes the argument an analytic function of x along the real axis. Since $kb\leqq2.5$, the first (and higher) derivative correction terms in equation 16 are small (less than 1 per cent) compared to $\psi(\phi)$ and can be ignored, since $\psi(\phi)$ is at best a minor part of $I(\phi)$ in equation 14. Hence,

$$\psi(\phi)=-2\pi kb\int_{4.5}^{\infty}\times\frac{\cos x\phi dx}{(x^2-k^2b^2)\left(\ln n_0-\ln x-i\dfrac{(kb)^{2x+1}}{\Gamma(2x+2)}\right)}\tag{17}$$

Next, it can readily be shown that the complex term in the integral of equation 17 can also be ignored. This yields

$$\psi(\phi)=-2\pi kb\int_{4.5}^{\infty}\frac{\cos x\phi dx}{(x^2-k^2b^2)(\ln n_0-\ln x)}\tag{18}$$

The integral in equation 18, which is to be interpreted in the principal-value sense, can be rewritten as follows

$$\psi(\phi)=\psi_1(\phi)+\psi_2(\phi)\tag{19A}$$

$$\psi_1(\phi)=-2\pi kb\int_{4.5}^{\infty}\frac{\cos x\phi}{\ln n_0-\ln x}\frac{dx}{x^2}\tag{19B}$$

$$\psi_2(\phi)=-2\pi kb\int_{4.5}^{\infty}\frac{\cos x\phi}{\ln n_0-\ln x}\times\frac{k^2b^2dx}{x^2(x^2-k^2b^2)}\tag{19C}$$

Since n_0 is quite large and $kb\leqq2.5$, 19(C) becomes, to a satisfactory approximation

$$\psi_2(\phi)\cong\frac{-2\pi kb}{\ln n_0-\ln 4.5}\int_{4.5}^{\infty}\frac{k^2b^2\cos x\phi}{x^2(x^2-k^2b^2)}dx$$

$$\cong\frac{-2\pi k^3b^3}{\ln\left(\dfrac{n_0}{4.5}\right)}\int_{4.5}^{\infty}\frac{\cos x\phi}{x^4}dx$$

$$=-\frac{2\pi}{\ln\left(\dfrac{n_0}{4.5}\right)}\left(\frac{kb}{4.5}\right)^3J_2(\phi)$$

This integral, $J_2(\phi)$, can be evaluated explicitly in terms of sines, cosines, and integral sines.

With the use of these results, an explicit equation for the current distribution can be written as

$$I(\phi)=\frac{V}{j\pi\zeta_0}\left\{\frac{1}{\alpha_0}+2\sum_1^4\frac{\cos n\phi}{\alpha_n}-\frac{2\pi}{\ln\left(\dfrac{n_0}{4.5}\right)}\times\left[\left(\frac{kb}{4.5}\right)J_1(\phi)+\left(\frac{kb}{4.5}\right)^3J_2(\phi)\right]\right\}\tag{20}$$

where

$$J_1(\phi)=\int_1^{\infty}\frac{\ln\left(\dfrac{n_0}{4.5}\right)}{\ln\left(\dfrac{n_0}{4.5}\right)-\ln x}\frac{\cos(4.5\phi)}{x^2}dx\tag{21A}$$

$$J_2(\phi)=\int_1^{\infty}\frac{\cos(4.5x\phi)}{x^4}dx\tag{21B}$$

$n_c=\dfrac{2b}{a}\epsilon^{-\gamma}$ and explicit equations for the α_n are given in Appendix III.

Note that the $J_k(\phi)$ integrals have only appreciable values near $\phi=0$. (An asymptotic equation for them, when $\phi>1$, can readily be obtained.) They cannot be approximated satisfactorily by a sinusoid and they are a partial explanation of why approximate methods of dealing with the loop antenna do not yield good quantitative results.

The equation for the impedance of the loop antenna becomes

Table II. Impedance of Loop Antennas

$\Omega=10,\ 2\pi b/a=148.41.\quad \Omega=11,\ 2\pi b/a=244.69$

R	X	$G\times10^3$	$B\times10^3$	kb	R	X	$G\times10^3$	$B\times10^3$
0.0051	62.59	0.0013	−15.98	0.05	0.0047	72.24	0.0009	−13.84
0.0410	128.0	0.0025	−7.812	0.10	0.0411	147.0	0.0019	−6.803
0.1577	203.7	0.0038	−4.908	0.15	0.1332	233.9	0.0028	−4.275
0.5936	297.7	0.0067	−3.360	0.20	0.5991	342.7	0.0051	−2.918
1.777	425.8	0.0098	−2.348	0.25	1.744	488.8	0.0073	−2.046
6.355	624.4	0.0163	−1.601	0.30	6.263	713.6	0.0123	−1.401
25.47	1003.1	0.0253	−0.9963	0.35	24.43	1136.6	0.0189	−0.8793
159.7	2063.4	0.0373	−0.4818	0.40	149.1	2294.6	0.0282	−0.4339
1679.4	−3205.9	0.0571	0.0109	0.45	2263.2	5631.6	0.0430	−0.0107
468.8	−2250.5	0.0887	0.4258	0.50	479.5	−2768.6	0.0607	0.3505
156.6	−1142.6	0.1177	0.8590	0.55	167.7	−1360.8	0.0892	0.7238
100.7	−756.0	0.1731	1.300	0.60	106.0	−891.3	0.1316	1.106
80.95	−555.7	0.2567	1.762	0.65	84.42	−650.5	0.1962	1.512
73.25	−430.4	0.3842	2.258	0.70	75.87	−500.7	0.2959	1.952
71.23	−341.9	0.5841	2.803	0.75	73.50	−394.9	0.4556	2.448
72.60	−274.3	0.9018	3.407	0.80	74.78	−314.0	0.7175	3.013
76.57	−219.5	1.416	4.061	0.85	78.92	−248.3	1.162	3.657
82.99	−173.3	2.248	4.694	0.90	85.67	−192.1	1.937	4.342
91.97	−132.5	3.534	5.093	0.95	95.37	−141.5	3.274	4.859
103.7	−95.58	5.214	4.808	1.00	108.1	−95.57	5.192	4.591
119.2	−61.60	6.621	3.422	1.05	125.5	−51.61	6.816	2.804
138.9	−29.61	6.884	1.468	1.10	148.3	−8.771	6.720	0.3974
164.8	−12.50	6.067	0.0046	1.15	179.4	33.22	5.389	−0.9979
199.2	25.90	4.936	−0.6415	1.20	222.4	73.70	4.051	−1.342
244.4	45.67	3.954	−0.7390	1.25	282.8	109.6	3.074	−1.192
302.2	53.12	3.210	−0.5642	1.30	367.3	132.4	2.410	−0.8682
371.0	37.06	2.669	−0.2666	1.35	480.3	121.7	1.956	−0.4955
438.4	−16.75	2.278	0.0870	1.40	604.5	50.75	1.643	−0.1379
475.9	−109.9	1.995	0.4607	1.45	677.1	−122.7	1.430	0.2592
455.1	−214.2	1.799	0.8465	1.50	627.6	−308.0	1.284	0.6302
384.9	−286.2	1.673	1.244	1.55	491.3	−412.2	1.195	1.002
302.8	−309.7	1.614	1.651	1.60	355.3	−425.4	1.156	1.385
234.4	−299.8	1.619	2.070	1.65	260.3	−393.3	1.170	1.768
185.2	−274.2	1.692	2.504	1.70	198.2	−346.3	1.245	2.175
153.5	−242.6	1.863	2.944	1.75	162.4	−295.3	1.430	2.600
133.7	−211.3	2.139	3.380	1.80	138.1	−253.9	1.653	3.040
122.6	−181.7	2.551	3.781	1.85	126.3	−213.3	2.055	3.471
118.1	−154.5	3.122	4.086	1.90	122.1	−176.4	2.658	3.832
118.5	−129.8	3.836	4.202	1.95	124.0	−142.6	3.471	3.994
123.2	−107.5	4.610	4.019	2.00	130.5	−111.5	4.429	3.786
131.6	−87.41	5.272	3.500	2.05	142.1	−82.63	5.259	3.058
143.9	−69.91	5.625	2.732	2.10	159.2	−56.28	5.583	1.974
159.8	−55.64	5.580	1.942	2.15	182.5	−33.00	5.307	0.9598
179.6	−45.83	5.228	1.335	2.20	212.8	−15.14	4.678	0.3328
202.5	−41.74	4.737	0.9764	2.25	250.7	−4.160	3.988	0.0662
227.6	−45.36	4.226	0.8424	2.30	295.8	−6.389	3.380	0.0730
251.3	−58.68	3.774	0.8812	2.35	342.5	−27.26	2.901	0.2309
269.6	−81.82	3.397	1.031	2.40	379.2	−70.53	2.535	0.4715
277.5	−112.5	3.094	1.255	2.45	397.6	−132.8	2.263	0.7559
271.3	−144.3	2.873	1.528	2.50	381.6	−196.0	2.074	1.065

Table III. Impedance of Loop Antennas

$\Omega=12,\ 2\pi b/a=403.43$

kb	R	X	$G\times10^3$	$B\times10^3$
0.05	0.0053	81.46	0.0008	−12.28
0.10	0.0419	167.0	0.0015	−5.986
0.15	0.1548	265.3	0.0022	−3.769
0.20	0.5967	386.2	0.0040	−2.589
0.25	1.721	549.5	0.0057	−1.820
0.30	6.042	797.4	0.0095	−1.254
0.35	23.50	1255.6	0.0149	−0.7962
0.40	131.5	2446.6	0.0219	−0.4076
0.45	1465.2	14,782.6	0.0337	−0.0340
0.50	588.8	−3477.0	0.0474	0.2799
0.55	185.6	−1617.7	0.0700	0.6102
0.60	106.1	−1011.2	0.1033	0.9842
0.65	89.32	−754.6	0.1547	1.307
0.70	79.48	−576.3	0.2348	1.703
0.75	76.52	−451.4	0.3650	2.153
0.80	77.55	−356.1	0.5839	2.681
0.85	81.63	−278.4	0.9698	3.308
0.90	88.50	−211.6	1.682	4.022
0.95	98.43	−151.5	3.017	4.642
1.00	112.0	−95.32	5.178	4.406
1.05	130.5	−41.05	6.974	−2.194
1.10	155.3	13.30	6.393	−0.5478
1.15	190.7	67.58	4.660	−1.652
1.20	239.3	125.4	3.278	−1.718
1.25	312.5	181.6	2.392	−1.390
1.30	422.0	227.9	1.834	−0.9905
1.35	584.2	236.5	1.471	−0.5954
1.40	788.1	142.3	1.229	−0.2218
1.45	926.1	−116.0	1.063	0.1332
1.50	837.1	−419.6	0.9546	0.4785
1.55	610.9	−559.7	0.8899	0.8153
1.60	414.5	−554.2	0.8654	1.157
1.65	287.7	−493.0	0.8829	1.513
1.70	212.5	−422.5	0.9503	1.889
1.75	168.8	−356.6	1.085	2.291
1.80	144.0	−297.6	1.317	2.722
1.85	131.3	−245.3	1.696	3.168
1.90	123.8	−193.8	2.295	3.594
1.95	129.7	−154.6	3.185	5.796
2.00	137.6	−113.9	4.314	3.571
2.05	151.7	−75.20	5.292	2.623
2.10	172.7	−38.28	5.518	1.223
2.15	195.8	−36.02	4.941	0.0909
2.20	243.0	26.49	4.067	−0.4433
2.25	257.2	48.43	3.277	−0.5340
2.30	367.0	53.74	2.668	−0.3907
2.35	447.1	28.71	2.227	−0.1430
2.40	520.0	−39.84	1.912	0.1465
2.45	552.7	−148.0	1.888	0.4522
2.50	521.0	−259.2	1.538	0.7841

$$Z=j\pi\zeta_0\left\{\frac{1}{\alpha_0}+2\sum_{1}^{4}\frac{1}{\alpha_n}-\frac{2\pi}{\ln\left(\frac{n_0}{4.5}\right)}\times\left[\left(\frac{kb}{4.5}\right)J_1(0)+\left(\frac{kb}{4.5}\right)^3J_3(0)\right]\right\}^{-1}\qquad(22)$$

This result, in connection with Appendix III, forms the basis for the impedance tables presented in Appendix I. The quantities $J_k(0)$ are explicitly given by the equation

$$J_1(0)=\frac{\ln\left(\frac{n_0}{4.5}\right)}{\left(\frac{n_0}{4.5}\right)}\int_{-\infty}^{\ln\left(\frac{n_0}{4.5}\right)}\frac{e^{+x}}{x}\,dx$$

$$J_3(0)=1/3$$

Results

The impedance of loop antennas for various values of b/a have been calculated with the use of equation 22. As a parameter, the quantity

$$\Omega=2\ln\frac{2\pi b}{a}\qquad(23)$$

has been chosen. Note that $2\pi b/a=c/a$, where c is the circumference of the antenna. Hence equation 23 represents a definition analogous to that used for dipole antennas.

In Appendix I, values of the impedance are tabulated for $0\leqq kb\leqq2.5$ and $\Omega=8$, 9, 10, 11, 12. They are also presented in graphic form. These impedances are useful for examining the operation of a loop antenna as a function of frequency. For laboratory purposes, however, it is sometimes convenient to have tables available appropriate to holding the frequency fixed and varying the size of the antenna. These tables, given in Appendix I, have been obtained by interpolation from the earlier tables.

It is perhaps worth while to comment on some of the more obvious features of these loop antenna impedances. As can be seen, the first antiresonance, occurring when the circumference of the loop approximates a half-wave-length, is extremely sharp. This well-known effect is easily explained by noting that a sufficiently small loop resembles closely a short-circuited quarter-wave-length transmission line and has a correspondingly sharp antiresonance.

Of equal interest is the rapid disappearance of resonances as the circumference of the antenna increases. Thus, for $\Omega\leqq9$, a second resonance point does not even exist. If one compares these impedances with those for a dipole antenna, it is seen that the two are similar, both qualitatively and quantitatively, for $c>\lambda$. The prime difference is that the loop is essentially more capacitive (by about 130 ohms) than a dipole. This can be explained on the basis that changed surfaces are closer together on a loop than on a dipole. This shift in reactance level by 130 ohms permits the dipole to have several resonances and antiresonances, whereas, as noted previously, a moderately thick loop ($\Omega<9$) has essentially

(*Continued on page 615*)

Table IV. Input Impedance for ka Constant

	a=3/16 Inch at λ=100 cm.		a=1/4 Inch at λ=100 Cm.		a=5/16 Inch at λ=100 Cm		
Kb	**R**	**X**	**R**	**X**	**R**	**X**	**Kb**
0.05							0.05
0.10							0.10
0.15		149		133			0.15
0.20		244		217		196	0.20
0.25	2	385	2	357		313	0.25
0.30	7	625	6	526	6	500	0.30
0.35	25	999	27	908	23	801	0.35
0.40	172	2209	167	1986	169	1871	0.40
0.45	19992	−3998	15202	−5153	13535	−3981	0.45
0.50	376	−2320	408	−2251	410	−2094	0.50
0.55	170	−1348	152	−1200	129	−963	0.55
0.60	104	−905	103	−820	103	−767	0.60
0.65	85	−670	83	−613	81	−570	0.65
0.70	77	−525	75	−486	74	−465	0.70
0.75	75	−422	73	−391	72	−367	0.75
0.80	76	−342	75	−317	74	−298	0.80
0.85	81	−270	80	−255	78	−242	0.85
0.90	89	−210	87	−197	86	−188	0.90
0.95	98	−155	97	−150	96	−142	0.95
1.00	112	−102	110	−95	108	−95	1.00
1.05	132	−39	129	−45	126	−50	1.05
1.10	157	20	154	6	151	−3	1.10
1.15	195Ω	81	189	60	184	45	1.15
1.20	245(12.447)	152	238	118	231	95	1.20
1.25			311	178	299	145	1.25
1.30			426	229	402	186	1.30
1.35			589	244	547	191	1.35
1.40			836	167	739	118	1.40
1.45			988	−108	853	−119	1.45
1.50			918	−448	800	−400	1.50
1.55			689Ω	−613	600	−547	1.55
1.60			447(12.446)	−623	418	−553	1.60
1.65					289	−497	1.65
1.70					215	−434	1.70
1.75					170	−365	1.75
1.80					146	−309	1.80
1.85					134	−252	1.85
1.90					126	−200	1.90
1.95					132	−160	1.95
2.00					141Ω	−113	2.00
2.05					157(12.494)	−72	2.05

Table V. Input Impedance of Loop Antenna for ka Constant

	a=3/8 Inch at λ=100 Cm.		a=1/2 Inch at λ=100 Cm.		a=3/4 Inch at λ=100 Cm		
Kb	**R**	**X**	**R**	**X**	**R**	**X**	**Kb**
0.05							0.05
0.10							0.10
0.15							0.15
0.20							0.20
0.25	2	278					0.25
0.30	6	454	6	416			0.30
0.35	24	768	22	666			0.35
0.40	171	1738	146	1456			0.40
0.45	1216	−3160	1052		6772	2646	0.45
0.50	400	−1959	431	−1821	513	−1660	0.50
0.55	153	−1053	157	−954	166	−837	0.55
0.60	100	−716	95	−653	103	−570	0.60
0.65	81	−542	80	−495	82	−435	0.65
0.70	73	−430	73	−390	73	−345	0.70
0.75	72	−348	71	−318	71	−280	0.75
0.80	73	−283	72	−261	72	−232	0.80
0.85	78	−230	76	−215	76	−192	0.85
0.90	84	−182	83	−170	82	−158	0.90
0.95	94	−138	92	−133	90	−125	0.95
1.00	106	−95	104	−95	101	−95	1.00
1.05	124	−53	122	−60	115	−67	1.05
1.10	148	−12	142	−23	134	−40	1.10
1.15	179	33	171	13	158	−11	1.15
1.20	224	77	212	48	192	13	1.20
1.25	287	120	268	83	234	33	1.25
1.30	379	150	343	98	291	41	1.30
1.35	511	151	448	94	361	30	1.35
1.40	661	80	560	35	434	−15	1.40
1.45	774	−123	632	−116	487	−112	1.45
1.50	719	−353	617	−295	476	−224	1.50
1.55	552	−497	492	−410	402	−307	1.55
1.60	389	−508	356	−428	316	−337	1.60
1.65	280	−460	264	−404	243	−327	1.65
1.70	210	−405	202	−360	191	−300	1.70
1.75	168	−343	164	−310	159	−265	1.75
1.80	143	−291	140	−266	136	−232	1.80
1.85	133	−241	128	−225	125	−200	1.85
1.90	124	−193	122	−182	118	−168	1.90
1.95	130	−155	126	−150	122	−138	1.95
2.00	138	−112	134	−111	128	−110	2.00
2.05	153	−75	147	−79	139	−85	2.05
2.10	181	−35	170	−45	156	−60	2.10
2.15	211	5	193	−15	179	−40	2.15
2.20	250	40	235	15	200	−18	2.20
2.25	310	65	286	34	248	−5	2.25
2.30	395	80	353	39	294	−8	2.30
2.35	489	53	427	16	343	−27	2.35
2.40	598	−25	499	−42	387	−70	2.40
2.45	653*	−174	541	−148	409	−136	2.45
2.50			517†	−262	398‡	−201	2.50

* (Ω12.486).

† (Ω11.951).

‡ (Ω11.141).

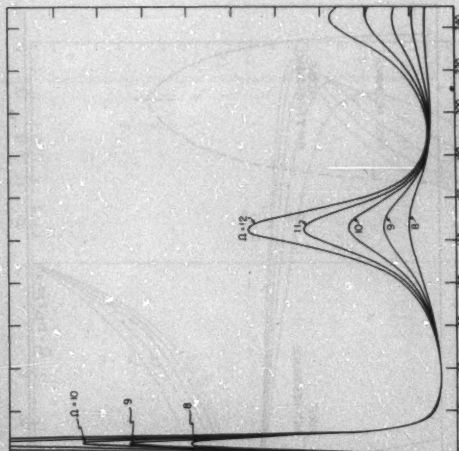


Fig. 5. Resistance of loop antennas as function of frequency

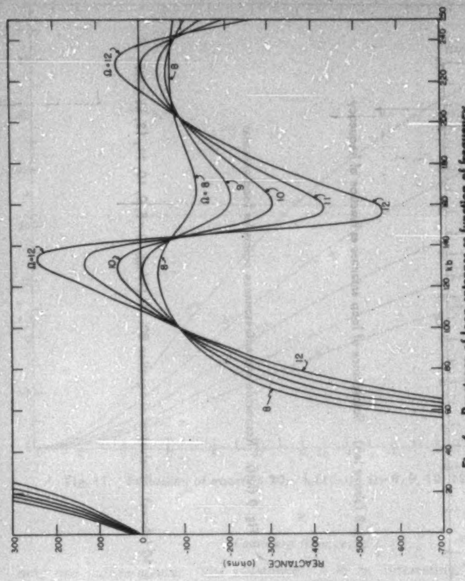


Fig. 6. Reactance of loop antennas as function of frequency

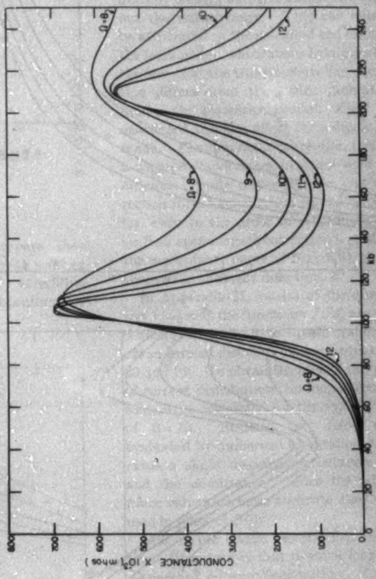
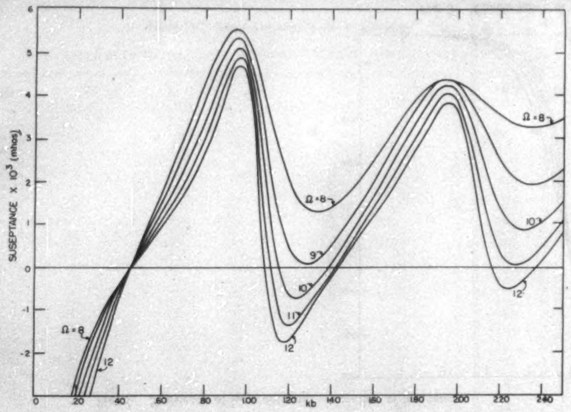
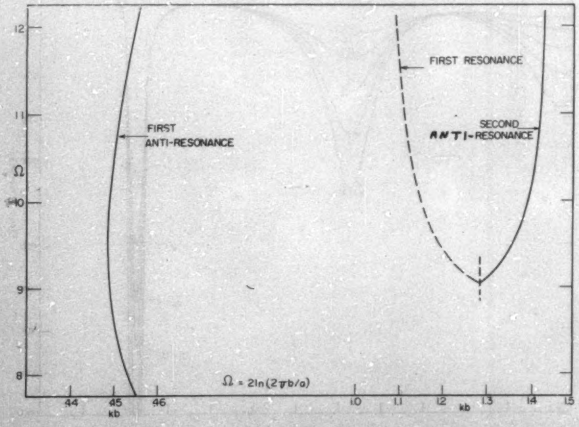


Fig. 7. Conductance of loop antennas as function of frequency





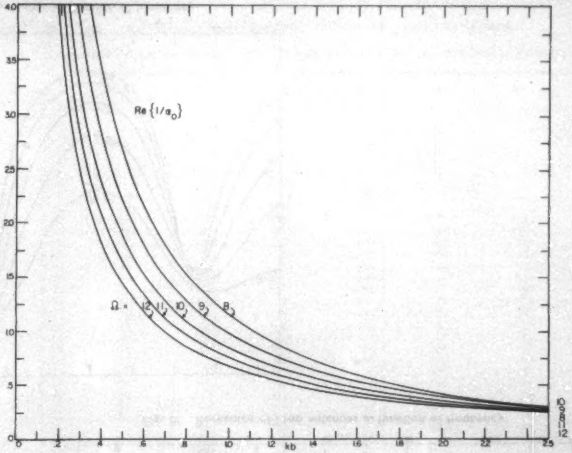


Fig. 8 (above left). Susceptance of loop antennas as function of frequency

Fig. 9 (left). Resonance and antiresonance points for loop antennas

Fig. 10 (above). Evaluation of equation 20: $R_e(1/\alpha_0)$; $\Omega=8, 9, 10, 11, 12$. $kb \leq 2.5$

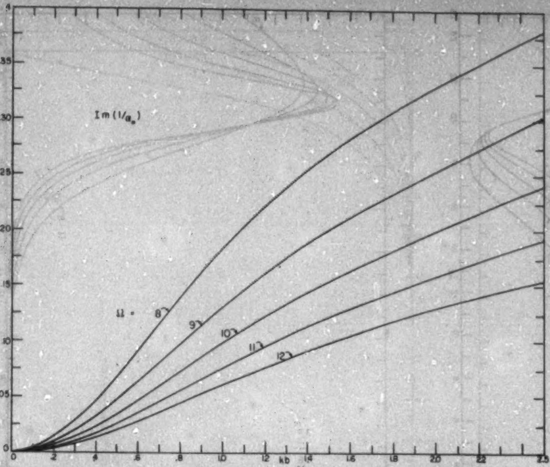


Fig. 11. Evaluation of equation 20: $I_m(1/\alpha_0)$; $\Omega=8, 9, 10, 11, 12$, $kb \leq 2.5$

(*Continued from page 611*)

only one antiresonance. The resistance curves for the loop and dipole are very similar, with the resistance minima having almost identical values.

It is interesting to compare these theoretical loop impedances with some experimentally measured ones. Miss Phyllis Kennedy of Cruft Laboratory has measured some loop impedances, using a half-loop over an image plane, and driven by a 2-wire line. The explicit configuration is indicated in Fig. 2(A). One set of the admittances measured by Miss Kennedy appears in Fig. 2(B) together with the corresponding theoretical curves. The agreement between the theoretical and experimental curves is seen to be excellent. It is seen that the resistance peaks near resonance on the theoretical conductance curves are slightly higher than those on the experimental curve. This could have been anticipated, as ohmic losses of the loop were not taken into account in the theoretical solution. The two susceptance curves differ by a slight additive amount throughout the entire range. This can readily be attributed so the so-called end effect of the feeding line, which arises from coupling the fact that the transmission-line excitation differs from the "slice generator" used in the theoretical model. King[4] has calculated this end effect for a dipole antenna. The dominant correction term is a negative capacitance in shunt with the antenna. Quite obviously, the end correction for a loop antenna should be similar, even to the order of magnitude. If such an approximate correction is made to the susceptance curve of Fig. 2(B), this is changed in the right direction.

In Appendix II, values of the quantities $1/\alpha_k$ and the functions $J_k(\phi)$ are presented graphically to facilitate evaluation of the current distribution using equations 20 and 21. To obtain an idea of the type of current distributions on loop antennas, some were calculated for the explicit case of $\Omega=10$. Because the $J_k(\phi)$ were evaluated by numerical integration, there exists a slight discrepancy between $I(0)$ and the admittance. Since the admittance values are more accurate, they were used in place of $I(0)$.

One of the classic assumptions in antenna literature is that a small loop has a constant current distribution. To examine the validity of the assumption, the actual current distributions were calculated for $\Omega=10$ and $kb=0.1$, 0.2, 0.3, and 0.4. These appear in Fig. 3. It is apparent that for the smallest loop, $kb=0.1$, the current varies in magnitude by about 5 per cent and hence can be considered reasonably constant. For $kb=0.2$, however, the variation is well over 10 per cent. On the basis of these results, one would be led to the conclusion that loops much larger than $kb=0.2$ cannot be considered small.

To obtain an idea of how the distribu-

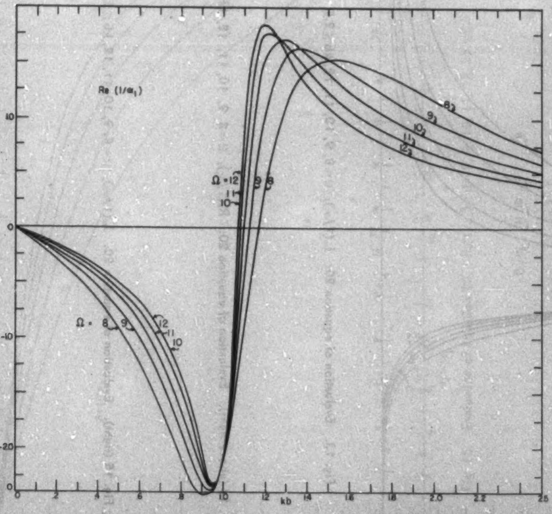


Fig. 12. Evaluation of equation 20: $R_e(1/\alpha_0)$; $\Omega=8, 9, 10, 11, 12$, $kb \leq 2.5$

(*Continued on page 618*)

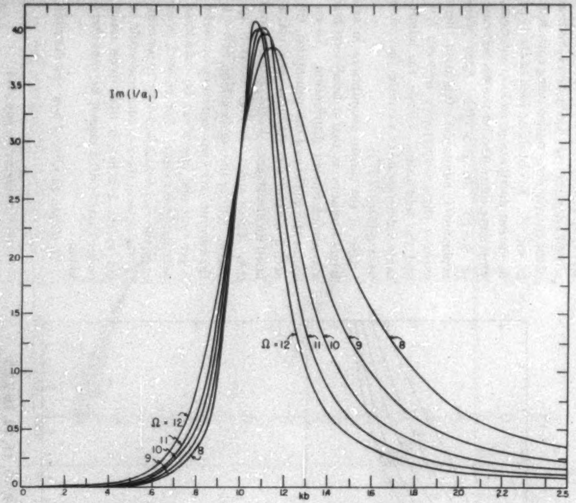


Fig. 13. Evaluation of equation 20: $I_m(1/\alpha_1)$; $\Omega=8, 9, 10, 11, 12$, $kb \leqq 2.5$

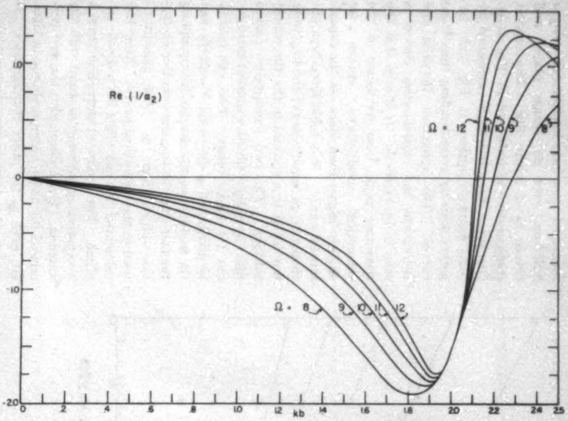


Fig. 14 (above right). Evaluation of equation 20: $R_e(1/\alpha_2)$; $\Omega=8, 9, 10, 11, 12$, $kb \leqq 2.5$

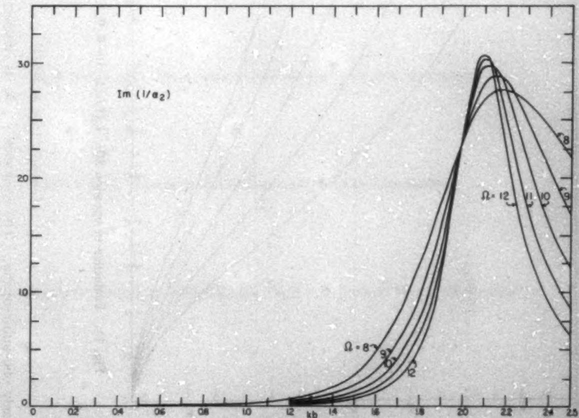
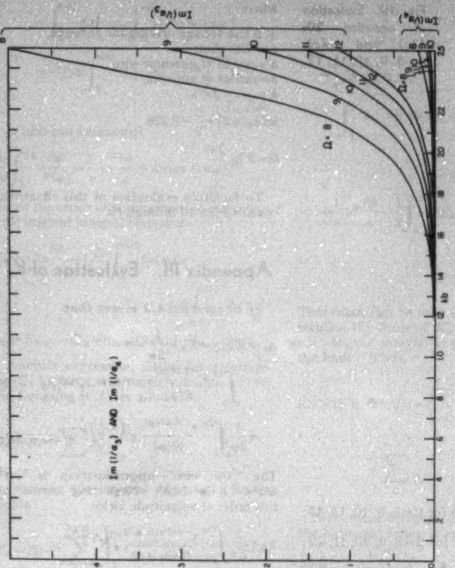
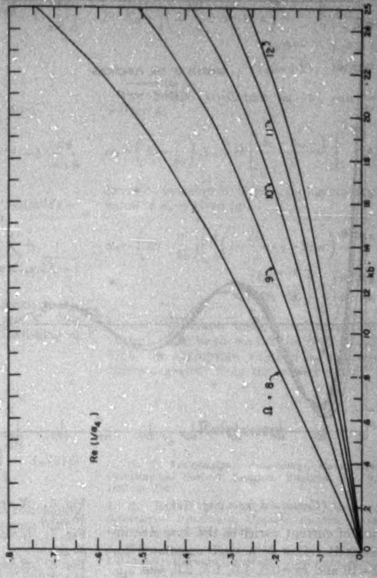


Fig. 15 (right). Evaluation of equation 20: $I_m(1/\alpha_2)$; $\Omega=8, 9, 10, 11, 12$, $kb \leqq 2.5$





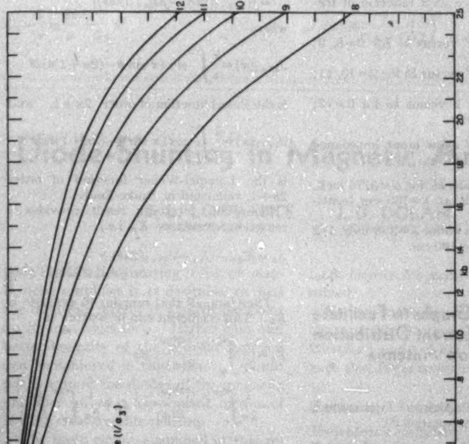


Fig. 16. Evaluation of equation 20: $R_e(1/\alpha_3)$; $\Omega = 8, 9, 10, 11, 12$, $kb \leqq 2.5$

Fig. 17 (above right). Evaluation of equation 20: $I_m(1/\alpha_3)$ and $I_m(1/\alpha_4)$; $\Omega = 8, 9, 10, 11, 12$, $kb \leqq 2.5$

Fig. 18 (right). Evaluation of equation 20: $R_e(1/\alpha_4)$; $\Omega = 8, 9, 10, 11, 12$, $kb \leqq 2.5$

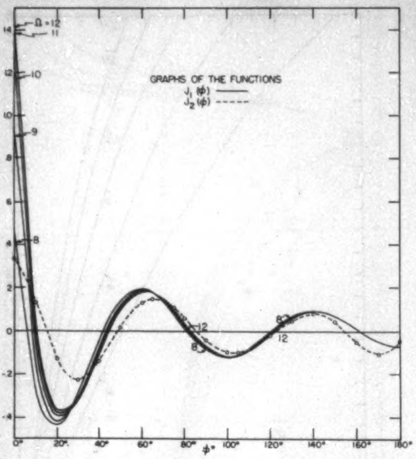


Fig. 19. Evaluation of equation 20: $J_1(\phi)$ and $J_2(\phi)$; $\Omega=8, 9, 10, 11, 12$, $0 \leq \phi \leq \pi$

(*Continued from page 615*)

tion of current varies as the loop size increased, values of it were calculated for $\Omega=10$ and $kb=0.5$, 1.0, 1.5, 2.0, and 2.5. These results appear in Fig. 4. Perhaps the most noticeable feature in these curves appears in the plots of magnitude and phase for the larger values kb. For values of $\phi<90$ degrees, it is apparent that the current distribution is beginning to approximate a traveling wave, in the sense that variations in the magnitude have been reduced and the phase is becoming linear. This is in agreement with the observation made in connection with the impedances, namely, that for larger kb the magnitude of the variation of the resistance is reduced.

Appendix I. Input Impedance of Loop Antennas

In Tables I through V impedances, $Z=R+jX$, are given in ohms and admittances, $Y=1/Z=G+jB$, are given in mhos. The loop radius is designated by b and the loop wire radius by a. The ratio b/a is expressed in terms of the parameter $\Omega=\ln 2\pi b/a$. Note that $2\pi b=c$, the circumference of the loop, and $kb=2\pi b/\lambda=c/\lambda$, where λ is the wave length. Thus kb is simply the circumference of the loop divided by the wave length.

The input impedance as a function of frequency is shown in graph form in Figs. 5 through 9 as follows:

Fig. 5: R versus kb for $\Omega=8, 9, 10, 11, 12$; $kb \leqq 2.5$.
Fig. 6: X versus kb for $\Omega=8, 9, 10, 11, 12$; $kb \leqq 2.5$.
Fig. 7: G versus kb for $\Omega=8, 9, 10, 11, 12$; $kb \leqq 2.5$.
Fig. 8: B versus kb for $\Omega=8, 9, 10, 11, 12$; $kb \leqq 2.5$.
Fig. 9: Locus of resonance and antiresonance points.

Tables I through III show input impedance and admittance as a function of frequency, as follows:

Table I: Z and Y versus kb for $\Omega=8, 9$; $kb \leqq 2.5$.
Table II: Z and Y versus kb for $\Omega=10, 11$; $kb \leqq 2.5$.
Table III: Z and Y versus kb for $\Omega=12$; $kb \leqq 2.5$.

Tables IV and V show input impedance for ka constant, as follows:

Table IV: Z versus kb for $a=3/16$ inch, 1/4 inch, 5/16 inch, at $\lambda=100$ cm (centimeters).
Table V: Z versus kb for $a=3/8$ inch, 1/2 inch, 3/4 inch, at $\lambda=100$ cm.

Appendix II. Graphs to Facilitate Evaluation of Current Distribution on a Loop Antenna

The current distribution on a loop antenna is given explicitly by equation 20

$$I(\phi)=\frac{V}{j\pi\zeta_0}\left\{\frac{1}{\alpha_0}+2\sum_{1}^{4}\frac{\cos n\phi}{\alpha_n}-\frac{2\pi}{\ln\frac{n_0}{4.5}}\times\left[\left(\frac{kb}{4.5}\right)J_1(\phi)+\left(\frac{kb}{4.5}\right)^3 J_2(\phi)\right]\right\}_{kb\leqq 2.5} \quad (20)$$

where

V is the voltage driving the antenna
$\zeta_0=120$ ohms
a = radius of antenna wire
b = radius of antenna
$k=\omega/c=2\pi/\lambda$

$$\ln(n_0/4.5)=\frac{\Omega}{2}-3.226$$

$$\Omega=2\ln\frac{2\pi b}{a}$$

To facilitate evaluation of this equation, consult Figs. 10 through 19.

Appendix III. Evaluation of K_n

From equation 4 it is seen that

$$\Delta_n=K_{n+1}-K_n=\frac{1}{2\pi}\times\int_0^{2\pi}\frac{\epsilon^{-jkbR(\phi)}}{R(\phi)}[\epsilon^{j(n+1)\phi}-\epsilon^{jn\phi}]d\phi \quad n>0$$

$$=\frac{1}{2\pi}\int_0^{2\pi}\frac{\epsilon^{-jkbR(\phi)}}{R(\phi)}\epsilon^{j\left(n+\frac{1}{2}\right)\phi}2j\sin\phi/2d\phi$$

The "thin-wire" approximation is that $k^2a^2<<1$, $a^2<<b^2$. Neglecting terms of this order of magnitude yields

$$\Delta_n=\frac{1}{2\pi}\int_0^{2\pi}\frac{\epsilon^{-jkb\sin\phi/2}}{2\sin\phi/2}\epsilon^{j\left(n+\frac{1}{2}\right)\phi}\times 2j\sin\phi/2d\phi+\text{term of order }(a^2/b^2)$$

$$=\frac{j}{\pi}\int_0^{\pi}\epsilon^{-2jkb\sin\theta+j(2n+1)\theta}d\theta$$

$$=j[J_{2n+1}(2kb)-j\Omega_{2n+1}(2kb)]$$

where

$$J_{2n+1}(x)=\frac{1}{\pi}\int_0^{\pi}\sin(x\sin\theta-(2n+1)\theta)d\theta$$

is the Bessel function of order $2n+1$, and

$$\Omega_{2n+1}(x)=\frac{1}{\pi}\int_0^{\pi}\cos(x\sin\theta-(2n+1)\theta)d\theta$$

is the Lommel-Weber function of order $2n+1$, tabulated in Janke-Emde.

Thus, the foregoing result provides a recursion formula for K_n, i.e.,

$$\Delta_n=K_{n+1}-K_n=\Omega_{2n+1}(2kb)+jJ_{2n+1}(2kb), \quad n>0$$

Therefore, all that remains to evaluate is K_0. This coefficient can be written as

$$K_0=\frac{1}{2\pi}\int_0^{2\pi}\frac{\epsilon^{-jkbR(\phi)}}{R(\phi)}d\phi$$

$$=\frac{1}{2\pi}\int_0^{2\pi}\frac{\epsilon^{-jkbR(\phi)}-1}{R(\phi)}d\phi+\frac{1}{2\pi}\int_0^{2\pi}\frac{d\phi}{R(\phi)}$$

Now

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{\epsilon^{-jkbR(\phi)}-1}{R(\phi)}d\phi$$

$$=\frac{1}{2\pi}\int_0^{2\pi}\frac{\epsilon^{-2jkb\sin\phi/2}-1}{2\sin\phi/2}d\phi+\text{term }0(a^2/b^2)$$

$$=\int_0^{2kb}dx\left(-\frac{j}{\pi}\int_0^{\pi}e^{-jx\sin\phi}d\phi\right)$$

$$=-1/2\int_0^{2kb}\Omega_0(x)dx-j/2\int_0^{2kb}\times J_0(x)dx$$

It also can be shown[3]

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{d\phi}{R(\phi)}=\frac{1}{\pi}\ln\frac{8b}{a}+\text{terms }0(a^2/b^2)$$

So, to the order of approximation consistent with original integral equation

$$K_0=\frac{1}{\pi}\ln\frac{8b}{a}-1/2\int_0^{2kb}\Omega_0(x)dx-j/2\int_0^{2kb}J_0(x)dx$$

$$\Delta_n=K_{n+1}-K_n=\Omega_{2n+1}(2kb)+jJ_{2n+1}(2kb)$$

Another expression, useful for determining K_n for large n, can also be found. From the foregoing it is seen that

$$K_n=K_0+\sum_0^{n-1}\Delta_n$$

Inserting the integral expressions for Δ_n yields

$$K_n=K_0+\frac{j}{\pi}\sum_0^{n-1}\int_0^{\pi}e^{-2jkb\sin\phi+j(2n+1)\phi}d\phi$$

$$=K_0+\frac{j}{\pi}\int_0^{\pi}e^{-2jkb\sin\phi}\left(\frac{e^{2jn\phi}-1}{\sin\phi}\right)d\phi$$

Inserting the value for K_0

$$K_n=\frac{1}{\pi}\ln\frac{8b}{a}+\frac{1}{2\pi}\times$$

$$\int_0^{\pi}(e^{-2jkb\sin\phi+2jn\phi}-1)\frac{d\phi}{\sin\phi}$$

$$=\frac{1}{\pi}\ln\frac{8b}{a}+\frac{1}{2\pi}\times$$

$$\int_0^{\pi}(e^{-2jkb\sin\phi}-1)\frac{e^{2jn\phi}}{\sin\phi}d\phi+\frac{1}{2\pi}\int_0^{\pi}(e^{2jn\phi}-1)\frac{d\phi}{\sin\phi}$$

$$=\frac{1}{\pi}\ln\frac{8b}{a}-\frac{1}{2}\int_0^{2kb}[\Omega_{2n}(x)+jJ_{2n}(x)]dx-\frac{2}{\pi}\sum_0^{n-1}\frac{1}{2K+1}$$

This result can be used conveniently to determine the form of K_n for large N. For $n>>kb$, the integral is small, vanishing in the limit. Thus

$$n>>1;\ n>kb,\ K_n\sim\left(\frac{1}{\pi}\ln\frac{8b}{a}-\frac{2}{\pi}\times\sum_0^{n-1}\frac{1}{2K+1}\right)-\frac{j}{2}\int_0^{2kb}J_{2n}(x)dx$$

Now, using Sterling's equation to evaluate the harmonic series, it can be shown that

$$\sum_0^{n-1}\frac{1}{2K+1}=\frac{\gamma}{2}+\frac{1}{2}\ln 4n,\ \gamma(=0.5772)\text{ is Eulers constant.}$$

Similarly, for $n>kb$

$$J_n(x)\cong\frac{1}{\Gamma(n+1)}\left(\frac{x}{2}\right)^n$$

So $K_n\sim\frac{1}{\pi}\left(\ln\frac{2b}{a}-\gamma-\ln n\right)-j\frac{(kb)^{n+1}}{\Gamma(2n+2)}\quad\begin{cases}n^2\gg 1\\ n>kb\end{cases}$

The Fourier coefficient α_n (4), can be written as

$$\alpha_n=\left(kb-\frac{n^2}{kb}\right)K_n+kb\left[\frac{\Delta_n-\Delta_n-1}{2}\right]$$

Since Δ_n vanishes for large n, the asymptotic value of α_n is given by

$$\alpha_n\sim\left(kb-\frac{n^2}{kb}\right)\left[\frac{1}{\pi}\left(\ln\frac{2b}{a}-\gamma-\ln n\right)-j\frac{(kb)^{2n+1}}{\Gamma(2n+2)}\right]\quad\begin{cases}n^2\gg 1\\ n>kb\end{cases}$$

Finally (by simple insertion of numerical values), it can be shown that for $k\leq 2.5$, and $n\geq 5$, the asymptotic value of α_n as given differs negligibly from the correct value.

References

1. **H. C. Pocklington.** *Proceedings*, Cambridge Philosophical Society, London, England, vol. 9, 1897, p. 324.

2. **E. Hallén.** *Nova Acta Regiae Societatis Scientiarum Upsaliensis*, Uppsala, Sweden, series IV, vol. 2, no. 4, Nov. 1938.

3. **Tung Chang.** *Technical Report no. 16*, Cruft Laboratory, Harvard University, Cambridge, Mass, July, 1947.

4. ANTENNA THEORY AND PRACTICE (book), **S. A. Schelkunoff, H. T. Friis.** John Wiley & Sons, Inc., New York, N. Y., section 13.12, 1952.

5. SUMMATION OF SLOWLY CONVERGING SERIES, **I. Gumowski.** *Journal of Applied Physics*, New York, N. Y., vol. 24, 1953, p. 1068.

6. THEORY OF ANTENNAS DRIVEN FROM TWO-WIRE LINE, **R. King.** *Ibid.*, vol. 20, 1949, p. 832.

Diode-Shunting in Magnetic Amplifiers

J. L. LOWRANCE
NONMEMBER AIEE

J. D. DOLAN
NONMEMBER AIEE

IN THE self-saturating type of magnetic amplifier it is desirable to bias the saturable reactor to a particular point on its hysteresis loop. Higher gain and better linearity of the transfer function can be achieved in this manner. Shunt paths around the diodes of the magnetic amplifier afford a convenient means of biasing the saturable reactors.

The basic circuit examined in this report is a conventional full-wave, push-pull magnetic amplifier. An analysis is made of the bias currents in two different circuit arrangements. The feedback effects which are found in this type of circuit are discussed and the effective feedback factors for each circuit are determined.

Various methods of obtaining bias for the saturable reactors are shown in Fig. 1. Circuits 1 and 2 of Fig. 1 offer the advantage that fewer components are required.

Terms and Symbols

B = feedback coefficient
E = supply voltage
$\hat{E}$ = peak value of supply voltage
$\bar{E}$ = average value of supply voltage
H_θ = magnetization force required to saturate the magnetic core at angle θ of half-cycle, oersteds
I_D = current through dummy load resistor after both cores have saturated
$\bar{I}_D$ = average current through dummy load resistor after both cores have saturated
I_L = load current
$\bar{I}_L$ = average value of load current over half-cycle
I_n = net control current seen by amplifier
I_s = current from signal source
K_e = open loop voltage gain
K_e' = closed loop voltage gain
K_I = open loop current gain
K_I' = closed loop current gain
l_m = effective magnetic path length of core, centimeters
N_c = control-winding turns
N_{fb} = feedback-winding turns
N_g = gate-winding turns
R_c = control-winding resistance
R_d = dynamic foreward resistance of diode
R_D = dummy load resistor

Paper 56-749, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 14, 1956; made available for printing May 8, 1956.

J. L. LOWRANCE and J. D. DOLAN are with the Bendix Aviation Corporation, South Bend, Ind.

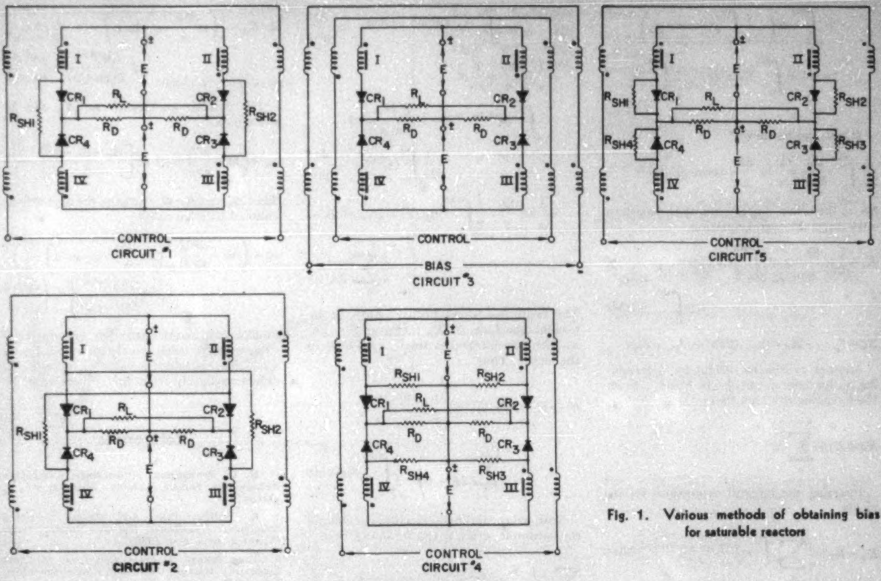


Fig. 1. Various methods of obtaining bias for saturable reactors

R_L = load resistor

R_s = signal source resistance

R_{sh} = shunt resistor

R_w = gate-winding resistance of core plus output resistance of power supply

V_{cr} = voltage across diode

$\bar{V}_{cr}$ = average voltage across diode over half-cycle

V_g = voltage absorbed by core with respect to gate winding

$\bar{V}_g$ = average value of voltage absorbed by core with respect to gate winding during half-cycle

V_{sh} = voltage across shunt resistor

$\bar{V}_{sh}$ = average voltage across shunt resistor over half-cycle

ΔV_{sh} = change in average value of voltage across shunt resistor over half cycle

Diode dynamic resistance = slope of voltage-current characteristic of diode after threshold voltage has been exceeded

Dummy load = resistors which are connected to power supply center tap from each pair of diodes

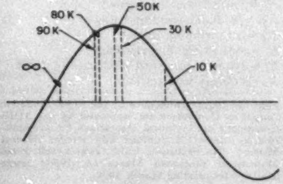


Fig. 2. Saturation angle versus shunt resistor

Gating half-cycle = interval during which supply voltage is positive with respect to anode of diode under consideration

Reset half-cycle = interval during which supply voltage is negative with respect to anode of diode under consideration

Shunt or leakage resistors = resistors which allow current to flow through gate winding during reset half-cycle

Gate winding = winding on saturable reactor which is connected to power supply

Explanation of Bias Techniques

Under conditions of no input signal the saturation angle of each core in the self-saturating type of magnetic amplifier is a function of the bias current that flows in its windings during the reset half-cycle.[1] This bias current can be supplied either by a separate winding on the core or by allowing current to flow in the reverse direction of the diode in the gate winding during the reset half-cycle. The latter method offers certain advantages because of ease in obtaining bias with little additional circuitry. This is the type of circuit considered in this report.

Under steady-state conditions, the voltage supported by a core averaged over a complete cycle is zero.

$$\int_0^{2\pi} e\,d\theta=0 \tag{1}$$

Assuming negligible excitation current, the voltage supported by the core during its gating half-cycle is equal to the integral of the supply voltage over the period before the core saturates.[2]

$$V_g=\frac{\hat{E}}{\pi}\int_0^{\theta} \sin\theta\, d\theta=\frac{\hat{E}}{\pi}(1-\cos\theta) \tag{2}$$

where $\hat{E}$ is the peak voltage and θ is the angle at which the core saturates, measured from zero with respect to the beginning of the half-cycle. Referring to equation 1, the sum of the voltages supported by the core averaged over the complete cycle is zero. Therefore the integral of the voltage absorbed by the core during the reset half-cycle can be expressed by equation 2.

$$V_g=\frac{\hat{E}}{\pi}(1-\cos\theta)$$

This may be written in terms of the average value of the supply voltage over the half-cycle.

$$V_g=\frac{E}{2}(1-\cos\theta) \tag{3}$$

As previously stated, for no-signal conditions, the angle at which the core saturates is a function of the bias current. In the case where shunt resistors are used, it is a function of the shunt resistance and the voltage across the shunt resistor.

The voltage across the shunt resistors will be analyzed in the time interval preceding and the time interval following the saturation of the cores. The first case to be considered will be for zero input signal. All cores will saturate at the same angle under this conditon.

Analysis of Shunt Voltage

Circuit 1 will be considered first.

Assume that cores 1 and 2 are positive with respect to the center tap of the power supply and cores 3 and 4 are negative with respect to this center tap.

Prior to saturation

$$\frac{E}{2}\int_0^\theta \sin\theta d\theta = V_{cr3}+V_{g3}=V_{cr4}+V_{g4} \tag{4}$$

After saturation

$$\frac{E}{2}\int_\theta^\pi \sin\theta d\theta = -I_DR_D+V_{cr3}+V_{g3}$$
$$= -I_DR_D+V_{cr4}+V_{g4} \tag{5}$$

The voltage absorbed by the core during the half-cycle is

$$\bar{V}_\theta=\frac{E}{2}(1-\cos\theta)$$

Therefore the average voltage across the diodes over the half-cycle is

$$\bar{V}_{cr3}=E+\bar{I}_DR_D-\frac{E}{2}(1-\cos\theta) \tag{6}$$

$$\bar{V}_{cr4}=E+\bar{I}_DR_D-\frac{E}{2}(1-\cos\theta) \tag{7}$$

$\bar{I}_D$ is the average value of the current flowing in the dummy load resistor of circuit 1 after both cores have saturated.

Adding the forward drop of the other diode being shunted by the shunt resistor during its gating half-cycle, the voltage across the shunt resistor becomes

$$\bar{V}_{sh1}=\frac{E}{2}(1+\cos\theta)+\bar{I}_DR_D+\bar{I}_DR_d \tag{8}$$

$$\bar{V}_{sh2}=\frac{E}{2}(1+\cos\theta)+\bar{I}_DR_D+\bar{I}_DR_d \tag{9}$$

R_d is the dynamic forward resistance of the diode.

The voltage across the dummy load resistor may be written in terms of the saturating angle θ, and the equation becomes

$$\frac{E}{2}(1+\cos\theta)=I_D(R_D+R_d+R_w)$$

$$\bar{V}_{sh1}=\bar{V}_{sh2}=\frac{E}{2}(1+\cos\theta)\times$$
$$\left(1+\frac{R_D+R_d}{R_D+R_d+R_w}\right) \tag{10}$$

For a given saturating angle and supply voltage, the value of reset magnetizing force can be calculated from the equation

$$H_\theta=\frac{0.4\pi NI}{lm} \tag{11}$$

where H_θ is the magnetizing force in oersteds and I is the magnetization current in amperes. In this case, I is the current through the shunt resistor.

The shunt resistance can be expressed in terms of the saturating angle of the core and the supply voltage

$$R_{sh1}=R_{sh2}=\frac{E(1+\cos\theta)}{\dfrac{H_\theta lm}{0.4\pi N_G}}\left(\frac{R_D+R_d+w/s}{R_D+R_d+R_w}\right) \tag{12}$$

The next step is to determine the relationship between the saturating angle and the magnetizing force. It is possible to obtain these data by empirical methods for a given core material and core geometry.[3] These data are then applicable to designs in which this type of core is used.

The change in saturation angle with various values of shunt resistors and constant supply voltage is shown in Fig. 2. These data were obtained by making measurements in the circuit and calculating the angle from the foregoing equations. Other techniques such as oscilloscope measurements and timing devices may be used to obtain the same data.[4]

An examination of circuit 2 shows that for zero signal conditions the equations for the voltage across the shunt resistors is the same as in circuit 1.

$$\bar{V}_{sh1}=\bar{V}_{sh2}=\frac{E}{2}(1+\cos\theta)+\bar{I}_DR_D+\bar{I}_DR_d \tag{13}$$

It may therefore be shown by deduction that for zero input signal conditions the two circuits are equivalent with respect to the bias circuit.

Analysis of Shunt Voltage with Input Signal

The preceding statement is true, however, only for zero input signal conditions. The circuit will now be considered for an input signal applied to the amplifier, over the linear region of the transfer function. The signal is assumed to be in sense which makes core 1 saturate before core 2. As in the previous cases, it is assumed that cores 1 and 2 are positive with respect to the power supply center tap.

CIRCUIT 1

Before core 1 saturates

$$\frac{E}{\pi}\int_0^{\theta_1}\sin\theta d\theta=V_{g1}=V_{g2}$$

$$\frac{E}{\pi}\int_0^{\theta_1}\sin\theta d\theta=V_{g3}+V_{cr3}=V_{g4}+V_{cr4}$$

After core 1 saturates but before core 2 saturates

$$\frac{E}{\pi}\int_{\theta_1}^{\theta_2}\sin\theta d\theta=V_{g1}+I_L(R_D+R_L)+$$
$$\left(\frac{2R_D+R_L}{R_D}\right)(R_w+R_d)I_L$$

$$\frac{E}{\pi}\int_{\theta_1}^{\theta_2}\sin\theta d\theta=V_{g2}+I_LR_D$$

$$\frac{E}{\pi}\int_{\theta_1}^{\theta_2}\sin\theta d\theta=V_{g3}+V_{cr3}-I_LR_D$$

$$\frac{E}{\pi}\int_{\theta_1}^{\theta_2}\sin\theta d\theta=V_{g4}+V_{cr4}-I_L(R_D+R_L)$$

After both cores have saturated

$$\frac{E}{\pi}\int_{\theta_2}^{\pi}\sin\theta d\theta=V_{g1}+I_D(R_w+R_d+R_D)$$

$$\frac{E}{\pi}\int_{\theta_2}^{\pi}\sin\theta d\theta=V_{g2}+I_D(R_w+R_d+R_D)$$

$$\frac{E}{\pi}\int_{\theta_2}^{\pi}\sin\theta d\theta=V_{g3}+V_{cr3}-I_DR_D$$

$$\frac{E}{\pi}\int_{\theta_2}^{\pi}\sin\theta d\theta=V_{g4}+V_{cr4}-I_DR_D$$

Adding the equations for the three time intervals together yields the equation for the average value of the supply voltage over the half-cycle.

$$\bar{E}=\bar{V}_{g1}+\bar{I}_L(R_D+R_L)+$$
$$\bar{I}_D(R_D+R_w+R_d)+$$
$$(R_w+R_d)\left(\frac{2R_D+R_L}{R_D}\right)\bar{I}_L \tag{14}$$

$$\bar{E}=\bar{V}_{g2}+\bar{I}_LR_D+\bar{I}_D(R_D+R_w+R_d) \tag{15}$$

$$\bar{E}=\bar{V}_{g3}+\bar{V}_{cr3}-\bar{I}_LR_D-\bar{I}_DR_D \tag{16}$$

$$\bar{E}=\bar{V}_{g4}+\bar{V}_{cr4}-\bar{I}_L(R_D+R_L)-\bar{I}_DR_D \tag{17}$$

Due to the input signal, the voltage absorbed by each core is no longer equal to the voltages absorbed by the other cores. For d-c output from the stage, cores 1 and 4 are wound in the same sense with respect to the control winding, and cores 2 and 3 are wound in the same sense, but opposite to cores 1 and 4. Therefore it may be evident that the effect of control signal upon cores 1 and 4 is equal and opposite to the effect upon cores 2 and 3.

Expressing the voltage absorbed by the core in terms of the saturating angles gives

$$\bar{E}=\frac{E}{2}(1-\cos\theta_1)+\bar{I}_L(R_D+R_L)+$$
$$\frac{2R_D+R_L}{R_D}(R_w+R_d)\bar{I}_L+$$
$$\bar{I}_D(R_D+R_d+R_w) \tag{18}$$

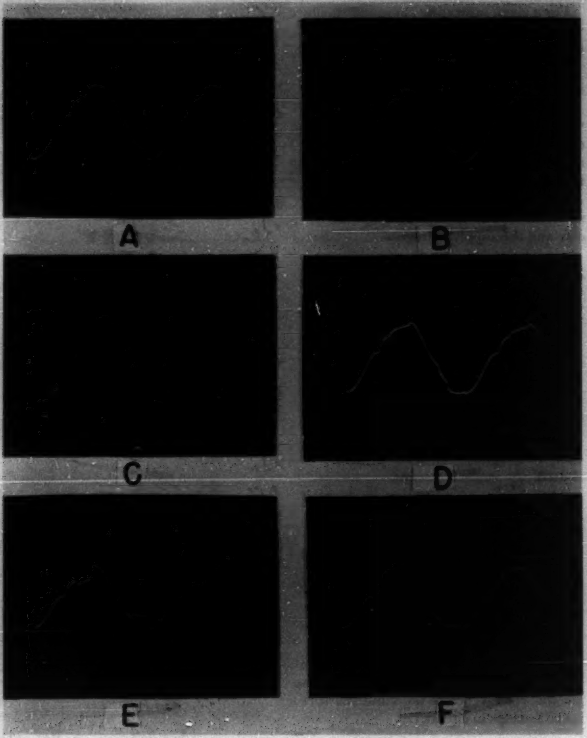


Fig. 3. Effect of control signal on shunt voltage

A—Circuit 1, voltage across R_{sh} for no-input signal
B—Circuit 2, voltage across R_{sh} for no-input signal
C—Circuit 1, voltage across R_{sh1} for an input signal positive with respect to dot
D—Circuit 2, voltage across R_{sh2} for an input signal positive with respect to the dot
E—Circuit 1, voltage across R_{sh2} for an input signal positive with respect to the dot
F—Circuit 2, voltage across R_{sh1} for an input signal positive with respect to the dot

$$\bar{E}=\frac{\bar{E}}{2}(1-\cos\theta_2)+\bar{I}_LR_D+\bar{I}_D(R_D+R_d+R_w) \tag{19}$$

$$\bar{E}=\frac{\bar{E}}{2}(1-\cos\theta_2)-\bar{I}_LR_D-\bar{I}_DR_D+\bar{V}_{cr3} \tag{20}$$

$$\bar{E}=\frac{\bar{E}}{2}(1-\cos\theta_1)-\bar{I}_L(R_L+R_D)-\bar{I}_DR_D+\bar{V}_{cr4} \tag{21}$$

Solving equations 18 and 19 for $\cos\theta_1$ and $\cos\theta_2$ and, substituting in equations 20 and 21, the voltages across diodes CR_3 and CR_4 are

$$\bar{V}_{cr3}=2\bar{I}_LR_D+\bar{I}_D(R_D+R_d+R_w) \tag{22}$$

$$\bar{V}_{cr4}=\bar{I}_L\left[\frac{2R_D+R_L}{R_D}(R_w+R_d)+2R_D+2R_L\right]+\bar{I}_D(R_D+R_d+R_w) \tag{23}$$

The voltages across the shunt resistors are

$$\bar{V}_{sh1}=\bar{V}_{cr4}+\bar{I}_LR_d\left(\frac{2R_D+R_L}{R_D}\right)+\bar{I}_DR_d \tag{24}$$

$$\bar{V}_{sh2}=\bar{V}_{cr3}+\bar{I}_DR_d \tag{25}$$

Substituting for the diode voltages from equations 22 and 23 gives

$$\bar{V}_{sh1}=\bar{I}_L\left[\frac{2R_D+R_L}{R_D}(R_w+2R_d)+2R_D+2R_L\right]+\bar{I}_D(R_D+R_w+2R_d) \tag{26}$$

$$\bar{V}_{sh2}=I_L(2R_D)+\bar{I}_L(R_D+R_w+2R_d) \tag{27}$$

From equations 26 and 27, it is seen that the voltage across the shunt resistor is a function of the load current and the current that flows in the dummy load, after both cores have saturated. Solving for the difference in the shunt voltages

$$\bar{V}_{sh1}-\bar{V}_{sh2}=\bar{I}_L\left[\frac{2R_D+R_L}{R_D}\times(R_w+2R_d)+2R_L\right] \tag{28}$$

From this equation, it is seen that the difference or unbalance in the shunt voltages is a function of the load current.

CIRCUIT 2

An examination of circuit 2 shows that the voltage appearing across the diodes may be expressed by the same equations derived for circuit 1. The voltages across the shunt resistors are

$$\bar{V}_{sh1}=-\bar{I}_LR_L+\bar{I}_DR_d+\bar{V}_{cr4} \tag{29}$$

$$\bar{V}_{sh2}=+\bar{I}_LR_L+R_d\left(\frac{2R_D+R_L}{R_D}\right)\times\bar{I}_L+R_d\bar{I}_D+\bar{V}_{cr3} \tag{30}$$

Substituting equations 22 and 23 for the diode voltages gives

$$\bar{V}_{sh1}=-\bar{I}_LR_L+\bar{I}_DR_d+\bar{I}_L\times\left[\frac{2R_D+R_L}{R_D}(R_w+R_d)+2R_D+2R_L\right]+\bar{I}_D(2R_D+R_w+R_d) \tag{31}$$

$$\bar{V}_{sh2}=+\bar{I}_LR_L+\bar{I}_DR_d+\bar{I}_L\frac{R_d}{R_D}(2R_D+R_L)+2\bar{I}_LR_D+\bar{I}_D(2R_D+R_w+R_d) \tag{32}$$

Subtracting equation 32 from 31 gives

$$\bar{V}_{sh1}-\bar{V}_{sh2}=\bar{I}_L\left(\frac{2R_D+R_L}{R_D}\right)(R_w) \tag{33}$$

As in the previous analysis of circuit 1, it is seen that the difference in shunt voltages is a function of the load current.

Explanation of Feedback Effect

It is assumed that the flux that is reset on the core during the reset half-cycle is a function of the average net ampere-turns that exist during the reset half-cycle. This is true for all practical purposes. The net ampere-turns are the algebraic sum of the control ampere-turns and the leakage ampere-turns. In the preceding mathematical analysis of the shunt voltage it is seen that a difference in shunt current exists which is a function of the amplifier output. A difference in the reset flux of cores 1 and 2

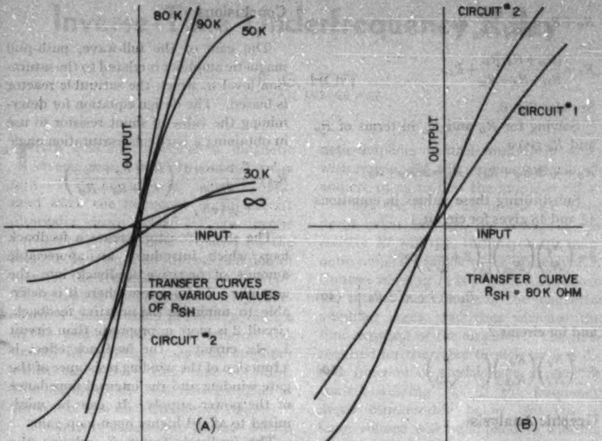


Fig. 4. Transfer curves. A—For various values of shunt resistance. B—For circuits 1 and 2

or cores 3 and 4 constitutes an output voltage from the amplifier.

On the basis of the statements in the preceding paragraph, the following equations are written

output = F_1 (input ampere-turns + Δ leakage current ampere-turns) (34)

Δ leakage current ampere-turns = F_2(output) (35)

Therefore

output = F_1, input + F_2 (output) (36)

This equation is of the form which commonly describes feedback systems.[5] Therefore it may be said that the change in shunt voltage as a function of output is a form of feedback.

Looking at the equation for the difference in shunt voltage in circuit 1, it is seen that V_{sh1} is larger than V_{sh2} for a positive signal with respect to the dot applied to the control winding. An increase in the shunt voltage V_{sh1} will tend to urge the cores 1 and 4 farther away from saturation. This is contrary to the effect of the control signal, which tends to urge the cores toward saturation. From this it may be deduced that the effect is that of negative feedback.

An examination of circuit 2 shows that for the same control circuit conditions V_{sh1} is larger than V_{sh2} for the half-cycle during which cores 3 and 4 are being reset. On the next half-cycle V_{sh2} will be larger than V_{sh1}. A quantitative measurement of the voltage across the shunts in circuit 2 would indicate that there is no difference in the shunt voltages. However, an examination of the wave form shows that the voltage is different on alternate half-cycles, as shown in Fig. 3(D) and (F). If V_{sh1} is larger than V_{sh2} during the interval that cores 3 and 4 are being reset, the change in leakage current will tend to urge the core farther away from saturation. This opposes the effect of the control signal and is therefore considered to be negative feedback.

In the next half-cycle, V_{sh1} is larger than V_{sh2}. In circuit 2 the leakage current through R_{sh2} is the bias current for core 1. The effect is the same as for the previous half-cycle, and is also considered to be negative feedback.

The difference in shunt voltages in circuit 1 may be divided into two parts

$$\bar{V}_{sh1}-\bar{V}_{sh2}=I_L(2R_w+4R_d)+\bar{V}_L\left(2+\frac{R_w+2R_d}{R_D}\right) \quad (37)$$

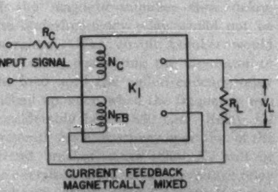


Fig. 5. Block diagram of amplifier with current feedback, magnetically mixed

From this equation it is seen that both load current and load voltage are factors in determining the feedback in the amplifier. Although the two are dependent upon each other, they are separate factors in the difference equation.

The difference in shunt voltages in circuit 2 may be arranged in a similar manner

$$\bar{V}_{sh1}-\bar{V}_{sh2}=I_L\left(2R_w+R_w\frac{R_L}{R_D}\right)=2R_wI_L+\frac{R_w}{R_D}\bar{V}_L \quad (38)$$

In this case, the feedback is a result of the impedance R_w and can be minimized by decreasing the winding resistance and the source impedance of the power supply.

In both circuits, there is feedback of voltage as a current and of current as a current. Fig. 4(B) shows the change in gain of a stage in circuits 1 and 2.

The feedback effect due to diode shunting can be related to the conventional feedback equations. This is equivalent to feedback into a separate winding, in the same winding configuration as the control winding. This type of circuit is shown in Fig. 5.[6] The gain expressions are, for open loop gain

$$K_I=\frac{I_L}{I_n}=\frac{I_L}{I_s}$$

I_s is the input signal and I_n is the net control current to which the amplifier responds.

Closed loop current gain is

$$K_I'=\frac{I_L}{I_s} \quad I_n=I_s-\frac{N_{fb}}{N_c}I_L$$

$$K_I'=\frac{I_L}{I_n+\frac{N_{fb}}{N_c}I_L}=\frac{K_I}{1+\frac{N_{fb}}{N_c}K_I} \quad (39)$$

Open loop voltage gain

$$K_e=K_I\frac{R_L}{R_s+R_c}$$

Closed loop voltage gain

$$K_e'=\frac{K_e}{1+\frac{N_{fb}}{N_c}\frac{R_s+R_c}{R_L}K_e} \quad (40)$$

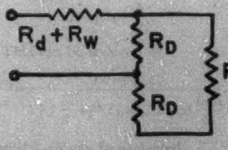


Fig. 6. Analysis of output network of amplifier with regard to maximum power transfer

In applying these equations to the shunt feedback effect, the gate winding is the counterpart of the feedback winding. The change in shunt voltage divided by shunt resistance is analogous to the feedback current.

Making these substitutions gives

$$K_I'=\frac{I_L}{I_n+\dfrac{N_g}{N_c}\dfrac{\Delta V_{sh}}{R_{sh}}} \qquad (41)$$

The differential change in shunt voltage is assumed to be equally divided

$$\Delta V_{sh}=\frac{1}{2}(V_{sh1}-V_{sh2}) \qquad (42)$$

Therefore for circuit 1

$$\Delta V_{sh}=\frac{I_L}{2}\left[\frac{2R_D+R_L}{R_D}(R_w+2R_d)+2R_L\right. \qquad (43)$$

$$K_I'=\frac{K_I}{1+\left(\dfrac{N_g}{N_c}\right)\left(\dfrac{1}{2R_{sh}}\right)\times\left[\dfrac{2R_D+R_L}{R_D}(R_w+2R_d)+2R_L\right]K_I} \qquad (44)$$

$$\beta=\left(\frac{N_g}{N_c}\right)\left(\frac{1}{2R_{sh}}\right)\times\left[\frac{2R_D+R_L}{R_D}(R_w+2R_d)+2R_L\right] \qquad (45)$$

For circuit 2

$$\Delta V_{sh}=\frac{I_L}{2}\left(\frac{2R_D+R_L}{R_D}\right)R_w \qquad (46)$$

$$K_I'=\frac{K_I}{1+\left(\dfrac{N_g}{N_c}\right)\left(\dfrac{1}{2R_{sh}}\right)\left(\dfrac{2R_D+R_L}{R_D}\right)R_wK_I} \qquad (47)$$

$$\beta=\frac{1}{2}\frac{N_g}{N_c}\frac{R_w}{R_{sh}}\left(\frac{2R_D+R_L}{R_D}\right) \qquad (48)$$

An analysis of the output network of the amplifier with regard to maximum power transfer yields the following equations

$$R_w+R_d=\frac{R_D+R_L}{2R_D+R_L}R_D$$

$$R_L=\frac{(R_w+R_d)R_D}{R_D+R_d+R_w}+R_D$$

See also Fig. 6.

Solving for R_D and R_L in terms of R_w and R_d gives

$$R_D=\sqrt{2}(R_w+R_d) \quad R_L=2(R_w+R_d)$$

Substituting these values in equations 45 and 48 gives for circuit 1

$$\beta=\left(\frac{N_g}{N_c}\right)\left(\frac{1}{R_{sh}}\right)\left[\left(3+\frac{1}{\sqrt{2}}\right)\times R_w+(4+\sqrt{2})R_d\right] \qquad (49)$$

and for circuit 2

$$\beta=\left(\frac{N_g}{N_c}\right)\left(\frac{R_w}{R_{sh}}\right)\left(1+\frac{1}{\sqrt{2}}\right) \qquad (50)$$

Graphic Analysis

The effect of the control signal on the shunt voltage is shown in Fig. 3.

For zero signal conditions, the shunt voltage in circuit 1 is equal to the shunt voltage in circuit 2. This is shown in Fig. 3(A) and (B). The sharp rise in the wave form occurs at the time the cores become saturated. Saturation occurs at approximately 70 degrees in this particular design.

Fig. 3(C) shows the increase in the voltage across R_{sh1} in circuit 1 when a signal voltage is applied which is positive with respect to the dot on the control winding. Fig. 3(E) shows the decrease in voltage across R_{sh2} for the same conditions.

Fig. 3(D) shows the increase in voltage across R_{sh1} in circuit 3, for a positive input voltage with respect to the dot of the control winding. On the following half-cycle the voltage decreases across the shunt. Examination of Fig. 3(F) indicates that this wave form is identical in form to that of Fig. 3(D) and out of phase with it by 1/2 cycle.

Conclusions

The gain of the full-wave, push-pull magnetic amplifier is related to the saturation level to which the saturable reactor is biased. The design equation for determining the value of shunt resistor to use in obtaining a particular saturation angle

$$R_{sh}=\frac{E(1+\cos\theta)}{\dfrac{H_\theta lm}{0.4\pi N_g}}\left(\frac{R_D+R_d+R_{wn}}{R_D+R_d+R_w}\right)$$

The shunt resistors create a feedback loop which introduces an appreciable amount of negative feedback into the amplifier. In instances where it is desirable to minimize the negative feedback, circuit 2 is more appropriate than circuit 1. In circuit 2, the feedback effect is a function of the winding resistance of the gate winding and the internal impedance of the power supply. It can be minimized to afford higher open-loop gain.

The feedback factor B, relating the feedback effect of the shunts on the open-loop gain of the amplifier, has been derived. Use of this factor is an aid in the analysis of the transfer function of this type of amplifier.

References

1. On the Mechanics of Magnetic Amplifier Operation, **R. A. Ramey.** *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

2. Self-Saturation in Magnetic Amplifiers, **W. J. Dornhoefer.** *Ibid.*, vol. 68, pt. II, 1949, pp. 835–50.

3. Techniques for Measuring Cascaded Self-Saturating Magnetic-Amplifier Performance, **Henry Kaplan, Gerald Wolff.** *Ibid.*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 581–84.

4. Oscillographic Techniques for the Evaluation of Magnetic Amplifier Response, **D. L. Critchlow.** *Ibid.*, vol. 74, pt. I, Nov. 1955, pp. 607–10.

5. Response of Physical Systems (book), **John D. Trimmer.** John Wiley & Sons, Inc., New York, N. Y., 1950, pp. 174–77.

6. Alteration of the Dynamic Response of Magnetic Amplifiers by Feedback, **R. O. Decker.** *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 658–65.

7. Fast Response with Magnetic Amplifiers, **D. G. Scorgie.** *Ibid.*, vol. 72, pt. I, 1953, pp. 741–49.

Vol.
Pt. 1
1956

Inverse-Time Underfrequency Relay

R. G. HOFT
ASSOCIATE MEMBER AIEE

THIS PAPER describes a magnetic inverse-time underfrequency relay unit. It contains a relay which is operated when the frequency of an aircraft alternator drops below 380 cps (cycles per second). The time required for the relay to pick up is an inverse function of how far the frequency is below 380 cps. The complete unit contains a frequency-sensitive circuit which provides a voltage signal proportional to alternator frequency for driving a magnetic amplifier which has a very long time constant. The output of the magnetic amplifier is the coil of the underfrequency relay.

The unit is designed to operate over a wide temperature range, from −55 C (degrees centigrade) to +71 C, and a wide voltage range, from 70 to 115 volts. Furthermore it meets military specifications for aircraft equipment. Figs. 1 and 2 show a complete unit.

Circuit Operation

The elementary diagram of the relay unit is shown in Fig. 3. The input transformer *T1* provides power input from a 115-volt 400-cps single-phase source. The 100-volt secondary is connected to the frequency-sensitive circuit consisting of inductor *L1* and capacitor *C1*. The output of the frequency-sensitive circuit is the voltage across resistor *R6*. The circuit is such that this voltage decreases with increasing frequency in the range of from 220 to 400 cps.

The voltage across resistor *R6* is rectified and filtered so that it appears as a relatively smooth d-c output across resistor *R1*. This voltage is then compared with a bias voltage across resistor *R2*. The bias voltage is the rectified and filtered output of the 20-volt transformer secondary. Resistor *R3* is provided to permit slightly greater than the voltage across resistor *R1*. This slight difference of voltage, when applied to the F_2 magnetic-amplifier control winding in series with resistor *R4*, will produce just enough ampere-turns to bias the amplifier off.

When the frequency of the a-c input drops below 380 cps, the voltage across resistor *R1* will rise, turning on the magnetic amplifier which picks up relay *K1*. Control winding F_1 is provided to reduce the rate of flux change in the magnetic-amplifier cores, and thus increase the time constant of the amplifier. The time required for the relay to pick up is a direct function of the voltage applied to control winding F_2. If the frequency drops considerably below 380 cps, a large voltage will be applied to winding F_2 in series with *R4* to turn on the amplistat quickly. When the input frequency is just slightly below 380 cps, only a small voltage will be applied to control winding F_2, and it will take a relatively long time for the magnetic amplifier to turn on.

Control winding F_3 provides positive feedback to give the magnetic amplifier a snap characteristic; it is either fully off or almost fully on. This snap action reduces the effect of variations in the pickup point of the relay, and increases the sensitivity of the circuit; a smaller change in frequency will turn the magnetic amplifier from off to on.

Relay *K2* is provided to prevent false operation during startup of the alternator. With reference to Fig. 3, the undervoltage relay *UV* must be energized, and the underfrequency relay *K1* must be de-energized in order to pick up the circuit breaker which connects the alternator to the load. Without relay *K2*, if the rate of rise in alternator voltage were much greater than the alternator acceleration during a startup, the undervoltage relay would pick up while the underfrequency relay remained out of operation. Because of the magnetic-amplifier time delay, the underfrequency relay would not be energized. The circuit breaker would then close, connecting the alternator to the load until the underfrequency relay picked up, even when the frequency was considerably below 380 cps. Relay *K2* prevents this momentary operation of the circuit breaker, since relay *K2* will be de-energized until the undervoltage relay picks up. While relay *K2* is out of operation, a large voltage from the frequency-sensitive circuit will be applied to control winding F_2. This voltage will cause relay *K1* to pick up quickly. Therefore, the underfrequency relay will always be picked up before relay *UV* is energized, so that the circuit breaker will not be operated until the frequency is correct permitting relay *K1* to drop out.

Nomenclature

E = average voltage, volts
E_C = average control voltage, volts
E_S = average supply voltage, volts
e = instantaneus voltage, volts
e_C = instantaneous control current voltage, volts
H_e = dynamic loop coercive force of core material
I_C = average control current, amperes
I_e = average core-exciting current, amperes
l_{fe} = length of magnetic oersteds circuit, centimeters
N_C = numbers of turns on control windings
R_C = control circuit resistance, ohms
T = period of supply frequency, seconds
T_d = time delay, seconds
t = time, seconds
Φ_M = maximum flux (crest value) produced by alternating supply voltage, maxwells
Φ_S = saturation flux, maxwells
φ = core flux, maxwells
τ_d = time delay, cycles

Circuit Design

FREQUENCY-SENSITIVE CIRCUIT

The objective in the design of this circuit was to obtain as much change in voltage output as possible for a given change in frequency. Several circuits were investigated, but the one shown in Fig. 3 was the most satisfactory. Resistor-capacitor circuits and saturable-reactor circuits did not provide as much voltage-output variation for a given frequency change as did this resonant circuit.

The basic portion of the frequency-sensitive circuit is shown in Fig. 4. The approximate transfer function of this network may be determined as follows:

$$\frac{E_{out}}{E_{in}}=\frac{\dfrac{R\left(pL_2+\dfrac{1}{pC}\right)}{R+pL_2+\dfrac{1}{pC}}}{pL_1+\dfrac{R\left(pL_2+\dfrac{1}{pC}\right)}{R+pL_2+\dfrac{1}{pC}}} \qquad (1)$$

This can be reduced to

$$\frac{E_{out}}{E_{in}}=\frac{(1+p^2L_2C)}{1+p\dfrac{L_1}{R}+p^2(L_1+L_2)C+p^3\dfrac{L_1L_2}{R}C} \qquad (2)$$

and if L_1/R can be neglected

$$\frac{E_{out}}{E_{in}}\approx\frac{(1+p^2L_2C)}{1+p^2(L_1+L_2)C} \qquad (3)$$

Paper 56-738, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 26, 1956; made available for printing May 4, 1956.

R. G. HOFT is with the General Electric Company, Schenectady, N. Y.

The basic circuit for the relay unit was proposed by H. H. Britten and D. L. Plette of the General Electric Company.

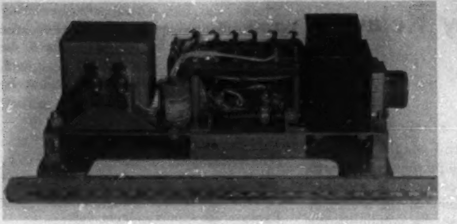


Fig. 1 (left). Inverse-time underfrequency relay, side view with cover removed

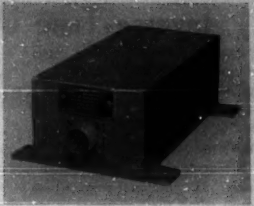


Fig. 2 (right). Inverse-time underfrequency relay unit, end view

This has an approximate frequency-response curve as shown in Fig. 5. Thus, in the range of from ω_1 to ω_2, the gain of the network will decrease as the frequency increases.

The gain of this network was calculated as a function of frequency by solving equation 3 using the following circuit constants:

L_1 = 0.635 henry
L_2 = 0.284 henry
C = 0.437 microfarad
R = 707 ohms

These values make ω_1 occur at approximately 1,580 radians per second, or 250 cps, and ω_2 at 2,840 radians per second, or 450 cps. The calculated curve is shown in Fig. 6. It is fairly linear over the range for the underfrequency relay.

The actual values of inductance, capacitance, and resistance selected were somewhat different from those indicated in the foregoing, so that it was possible to use a standard capacitor size. For the actual values shown in Fig. 3

$$\omega_1=\frac{1}{(L_1+L_2)C}=\frac{1}{0.8\times 0.5\times 10^{-6}}=1,580 \text{ radians per second, or 250 cps}$$

$$\omega_2=\frac{1}{L_2C}=\frac{1}{0.25\times 0.5\times 10^{-6}}=2,910 \text{ radians per second, or 460 cps}$$

The value of resistor *R6* is not critical, and, in fact, it can be eliminated without appreciably changing the output-voltage-versus-frequency curve from 220 to 380 cps. However, it was included in order to provide enough damping in the circuit so that the voltage applied to rectifiers *CR1* through *CR4* would not become too high at the resonant frequencies. The maximum rms voltage applied to these rectifiers was approximately 100 volts,

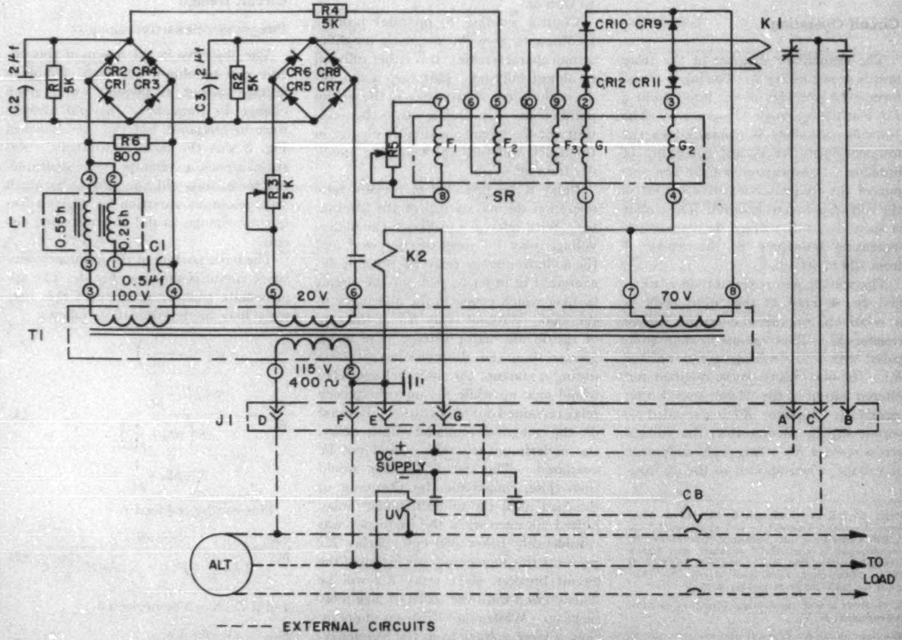

R4
5K
CR10 CR9
K1
2 μf
C2
R1
5K
CR2 CR4
CR1 CR3
2 μf
C3
R2
5K
CR6 CR8
CR5 CR7
CR12 CR11
R6
800
R5
F1
F2
F3
G1
G2
SR
R3
5K
L1
0.55h
0.25h
C1
0.5μf
100 V
K2
20 V
70 V
T1
115 V
400 ~
J1
D
F
E
G
A
C
B
DC SUPPLY
+
−
UV
CB
ALT
TO LOAD
---- EXTERNAL CIRCUITS

Fig. 3. Schematic diagram of inverse-time underfrequency relay

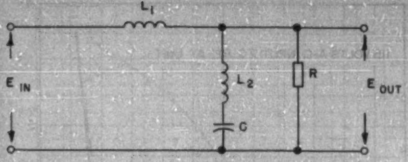


Fig. 4. Frequency-sensitive circuit

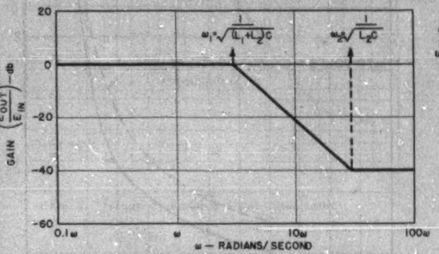


Fig. 5. Approximate frequency response curve of frequency-sensitive network

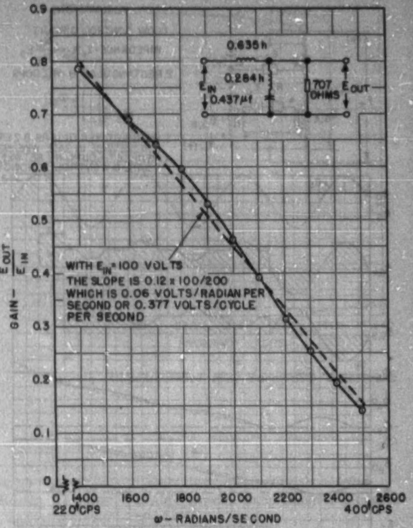


Fig. 6. Calculated gain versus frequency curve

with resistor *R6* equal to 800 ohms.

The filter capacitors *C2* and *C3* were extremely important. With the magnetic-amplifier supply voltage disconnected, the frequency-sensitive circuit would operate fairly well without these filter capacitors. The gain of the circuit was reduced somewhat because of ripple and the phase shift of the output voltages across the two bridge rectifiers, but otherwise the circuit operated properly. When the supply voltage was turned on, a change in frequency would not produce nearly as much effect on the magnetic-amplifier output as was expected. This was found to be caused by the induced current in control winding F_2 added to the ripple in the frequency-sensitive circuit rectifier outputs. The filtering is such that the frequency-sensitive circuit has approximately the effect on the amplifier that it should have based on the magnetic amplifier and the frequency-sensitive circuit gains separately operated.

MAGNETIC AMPLIFIER CIRCUIT

The magnetic amplifier used is a conventional, single-phase bridge-type amplifier. The supply voltage was determined by the requirements of relay *K1*. At the minimum input voltage to the relay unit of 70 volts, the magnetic-amplifier output current must be sufficient to pick up the relay. At the maximum input voltage of 115 volts, the maximum output current must not damage the relay. Both of these requirements were satisfied by a maximum magnetic-amplifier current output of approximately 60 ma (milliamperes) average and a maximum supply voltage to the magnetic amplifier of 70 volts. The maximum pickup current for relay *K1* is 20 ma and the largest permissible continuous coil current is 60 ma.

The magnetic amplifier was designed considerably underexcited; i.e., the normal supply of 70 volts at 380 cps produces a core flux amplitude Φ_M of approximately 20 per cent of Φ_S, provided that the cores are not driven into saturation. Arnold Engineering Company Deltamax cores are used with a saturation flux density B_S of 90,000 lines per square inch. The actual amplitude of the flux density for a supply of 70 volts and 380 cps is 18,400 lines per square inch for the cores which are used.

This underexcitation produces the required inverse-time characteristic, relatively independently of supply-voltage variation. When a step change in the d-c control voltage is applied to an underexcited magnetic amplifier which is biased off initially, it takes a period of time before the amplifier cores will saturate. The time before the cores saturate is a function of the instantaneous flux level when the step is applied, and of the magnitude of the step voltage. The steady-state change in flux level is produced by the d-c volt-seconds applied to the control circuit.

The response of an underexcited single-phase bridge magnetic amplifier to a step change in control voltage is shown in Fig. 7. The magnitude of the control-voltage step in this illustration was arbitrarily selected to be equal to $2N_C/N_G E_S$. Thus, the d-c control voltage referred to the gate circuits is $4/\pi$ times the peak value of the supply voltage. The d-c control volt-seconds are then equal to twice the a-c supply volt-seconds on each half-cycle.

It was assumed that the magnetic amplifier was off prior to the application of the step, and that the core fluxes were just reaching negative saturation on alternate half-cycles. This would be realized physically by having a slight but sufficient negative bias on the amplifier to drive the core fluxes slightly into negative saturation on alternate half-cycles. The core flux swings in the positive direction are produced by the alternating voltage applied to the gate circuits, and the negative flux swings are produced by the alternating voltage induced in the control-winding circuit. Prior to the application of the step of

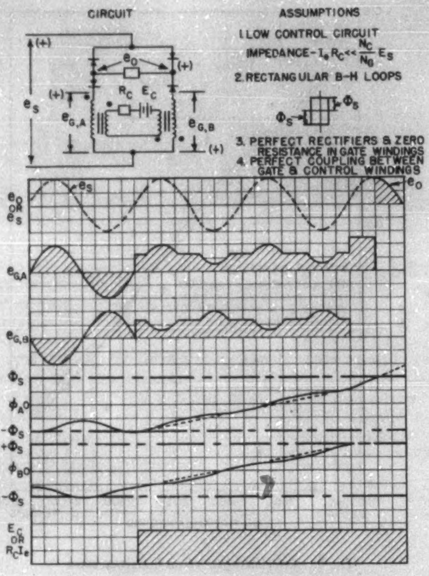


Fig. 7. Response to step change in control voltage

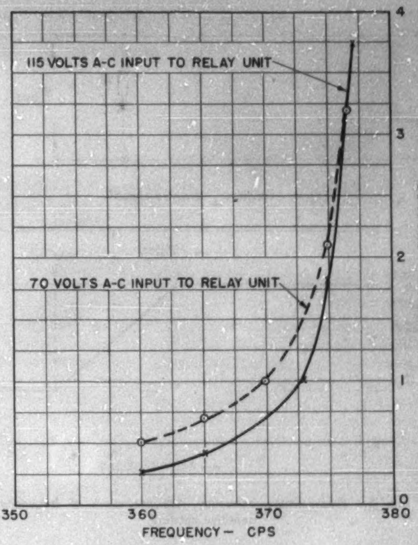


Fig. 8. Time–frequency curve

d-c control voltage, the control voltage is

$$E_C = R_C I_C = R_C I_e = R_C \frac{H_c l_{fe}}{0.4\pi N_C} \tag{4}$$

However this is very small, and has been neglected on the control-voltage curve in Fig. 7.

The wave shapes in Fig. 7 show how the core fluxes swing gradually toward positive saturation on consecutive half-cycles after the control-voltage step has been applied. During the first part of each half-cycle, one half of the d-c control voltage appears on each core. This control voltage reflected into the gate circuit is the only gate voltage on each core during this period. The rectifier cannot conduct in the gate circuit which is receiving a positive alternating voltage until the a-c gate voltage exceeds the reflected d-c control voltage. When this occurs, the a-c supply voltage becomes the gate voltage on the core whose rectifier is conducting. The gate voltage on the second core is the reflected total d-c control E_C minus the reflected alternating voltage induced in the control circuit from the conducting core. At the end of each half-cycle when the alternating voltage in the conducting gate drops below one half of the reflected d-c control voltage, again the gate voltages are each one half of the d-c control voltage.

During each complete cycle, the fluxes in each core move toward positive saturation by an amount equal to one half of the d-c control volts-seconds for a complete cycle. The net a-c flux swing in either core over a complete cycle is zero. Therefore, the magnitude of the d-c control-voltage step determines the time required for the core fluxes to reach positive saturation after the step has been applied. Since no magnetic-amplifier output can result until the cores saturate, there is a definite delay between the instant that a positive control-voltage step is applied and an output voltage is produced. (See Fig. 7.) This delay can be determined analytically as follows. In general

$$\frac{d\varphi}{dt} = \frac{e_C 10^8}{N_C} \tag{5}$$

Since there are two cores, the differential change in core flux in core A or core B, caused by the d-c control voltage E_C, when resistance drops are neglected, is

$$d\varphi_A = d\varphi_C = \frac{E_C 10^8}{2N_C} dt \tag{6}$$

The change in core flux on each cycle is exactly equivalent to having one half of the d-c control voltage on each core for the complete cycle, as shown in Fig. 7. Thus

$$\Delta\varphi_A/\text{cycle} = \Delta\varphi_B/\text{cycle} = \frac{E_C 10^8}{2N_C} T \tag{7}$$

Now, prior to the application of the step change in control voltage, the core fluxes alternate about a flux level Φ_M above negative saturation. The total flux change $\Delta\varphi_T$ required before the cores can reach saturation, the alternating component of the fluxes being neglected, is then

$$\Delta\varphi_T = 2\Phi_S - \Phi_M \tag{8}$$

If T_d is the time delay between when the step change in control voltage occurs and when the magnetic amplifier is fully on, the time delay, in cycles, of the magnetic-amplifier supply frequency is

$$\tau_d \approx \frac{2\Phi_S - \Phi_M}{\frac{E_C 10^8}{2N_C} T} \tag{9}$$

or, in seconds

$$T_d \approx \frac{2N_C}{E_C 10^8}(2\Phi_S - \Phi_M) \tag{10}$$

and therefore T_d is inversely proportional to the d-c control voltage.

For the conditions shown in Fig. 7, the time delay is

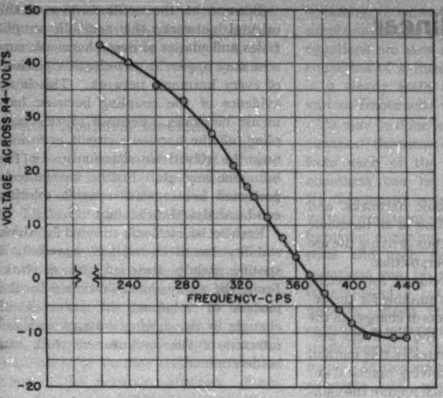


Fig. 9. Frequency-sensitive circuit characteristic

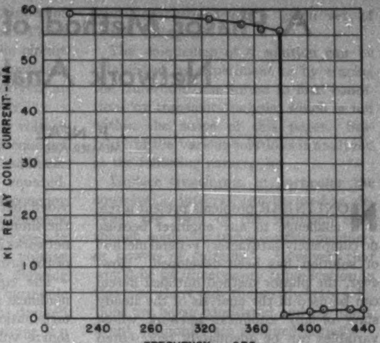


Fig. 10. Underfrequency relay steady-state characteristic

$$\tau_d \approx \frac{2\Phi_s - 10\% 2\Phi_s}{40\% 2\Phi_s/\text{cycle}} = 2.25 \text{ cycles} \qquad (11)$$

The expression for T_d in equation 9 or 10 is not exact because it does not consider when, in the cycle, the control-voltage step is applied, nor does it consider the time delay between when the cores first begin to saturate and when the magnetic amplifier is fully on. However, if the time delay is considerably greater than the period of the supply frequency, and if the amplifier goes from partially on to fully on in a few cycles, equation 9 or 10 will be an accurate measure of the magnetic-amplifier time delay.

It will be noted from the wave shapes in Fig. 7 that if the a-c flux swing is much less than $2\Phi_s$, the time delay will be relatively independent of alternating voltage. The effect of the control winding shunted by resistor *R5* (Fig. 3) is to increase the time delay. This may be considered as essentially placing a shunting resistor across control winding F_2, which reduces the control voltage by the ratio of this reflected resistance to resistor *R4* plus the reflected resistance.

Test Data

Figs. 8, 9, and 10 show test data for a sample unit. The test procedure for obtaining this data is explained in the Appendix. Fig. 8 is a plot of the measured time-frequency curve for the relay unit. Each point on this curve shows the time delay before the *K1* relay picked up after instantaneously dropping the input frequency from 380 cps to the frequency for the particular point on the curve. With 115 a-c input volts the *K1* relay dropped out at 386 cps, and with 70 a-c input volts the relay dropped out at 387 cps. Fig. 9 shows an output voltage versus frequency curve for the frequency-sensitive portion of the circuit; Fig. 10 is a plot of the *K1* relay current versus input frequency to the relay unit.

Conclusions

This magnetic-amplifier inverse-time underfrequency relay is a simple and reliable unit for indicating underfrequency on an a-c bus. Relatively long time delays are possible with the underexcited magnetic-amplifier circuit. This circuit combined with the frequency-sensitive circuit provides a fairly accurate inverse-time characteristic. Although no temperature tests were conducted, resistors with low-temperature coefficients and a special, mineral-oil-filled 0.5-microfarad capacitor were used, so that the circuit should not be appreciably affected by temperature variations.

Appendix. Test Procedure

The underfrequency relay unit is somewhat difficult to test with standard generating equipment. It is necessary to change the frequency of the a-c supply very quickly, and this is difficult with a motor-generator set. One method, however, is to use an electronic oscillator and an audio-amplifier as the a-c power supply. This was used in the General Electric Company laboratory, to test the development samples of the relay unit and adjust them for proper operation.

The following test equipment was required:

1 Kron-Hite oscillator model no. *440-A*

1 Altec Lansing audio amplifier catalogue no. *A256C*

1 Berkeley counter model no. *5500E*

The test procedure listed below was followed to test and adjust the relay units:

1. Connect the relay unit to the audio-amplier oscillator power supply.
2. Set the a-c input at 115 volts and 377 cps.
3. Adjust the bias resistor *R3* until the relay just picks up; this can be indicated by connecting an ohmmeter across terminals *A* and *C* on the unit.
4. Connect the Berkeley counter to record the time between when the oscillator frequency is changed and when the relay picks up. (This was done by connecting the starting input of the counter to a 1.5-volt battery through a push button. This button was pressed at the same time that the oscillator push buttons were pressed to change the input frequency. The "stop" input to the counter was connected to a battery in series with the normally open *K1* relay contacts.)
5. Press the oscillator buttons to change the input frequency and the auxiliary push button simultaneously, and record the time indicated on the counter.
6. Repeat this for several different frequency steps down from 380 cps.
7. If the time frequency curve is not as required, adjust *R5* to change the time delay, and repeat steps 5 and 6 until the desired curve is obtained.

Reference

1. Magnetic Amplifiers (book), Dr. H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955.

A Phasor Method of Nonlinear Network Analysis

J. P. NEAL
MEMBER AIEE

NONLINEAR problems present a real challenge to any engineer because no completely satisfactory general method of solution has been developed. However, the phasor method outlined herein can be used in the analysis of the steady state of any nonlinear network whose variables can be adequately represented by the sum of a limited number of simple sinusoidal functions of time. The method is the result of an effort to digest some of the classical approaches to the solution of nonlinear problems so they can be applied by means of elementary mathematics and at the same time provide a grasp of the influence of the individual nonlinear elements on the steady state of the nonlinear networks. It has the object of providing a tool for nonlinear analysis and synthesis to engineers who have not had extensive training in that area.

The direct attack used in this method requires some engineering judgment for the determination of the frequencies which are significant in the analysis. However, in practical engineering problems, such as the one which is referred to in the illustration of this method, adequate engineering information is often obtainable from experiment and preliminary analysis of the possible resonant frequencies of the physical system represented by the particular nonlinear network involved.

It is well known that the principle of superposition does not apply to nonlinear systems. This is often stated as an objection to employing the principle of equivalent linearization in nonlinear solutions. However, the method of this paper does not treat each harmonic independently. Mutual coupling is provided between harmonic networks, and the simultaneous solution of all harmonic networks is required, thus denying the use of the principle of superposition.

The equations developed for specific nonlinear solutions are suitable for the determination of the effects of changes in the source voltages or in the nonlinear network parameters. Therefore, this method provides a means for solving practical engineering problems which require the elimination of undesirable effects or the creation of desirable effects caused by the network nonlinearities.

This paper is presented in terms of the mesh method with source functions which are regulated voltages, and nonlinear parameters which are functions of currents. However, where source functions are regulated currents and nonlinear parameters are functions of voltages, the node method may be similarly employed.

Summary of the Method

The phasor method of nonlinear network analysis may be summarized in the following manner. Consider an electrical nonlinear network where all sources are regulated voltages and can be represented by finite sums of simple sinusoidal functions of time. The frequencies significant to the performance of the network are initially selected on the basis of experience. Mesh currents are then established, in symbolic form, with harmonic components at the significant frequencies. By substituting the symbolic forms of the mesh currents and simplifying the resultant expressions by trigonometric identities, each element voltage is arranged as the sum of a finite series of simple sinusoidal functions of time.

A set of circuit equations for each significant harmonic frequency is finally obtained by the substitution of the corresponding frequency term of each element voltage for each appearance of that voltage in the circuit (voltage) equations of the nonlinear network. Now, a harmonic network can be defined for each significant frequency by the corresponding harmonic set of circuit equations. Also, the harmonic voltages and currents can be represented in phasor form.

Because of the nonlinearities of the original network, the symbolic amplitudes and phases of every harmonic current may appear in the circuit equations of every harmonic network. This is an evidence of the coupling between harmonic networks. Because of this coupling, all the harmonic circuit equations must be solved simultaneously for the harmonic currents. Finally, by the tests presented herein, each specific solution can be analyzed for stability.

Each set of harmonic currents for which stability has been verified represents a specific steady state of the network. From the expressions developed in the process of solution, the influence of changes in the driving voltages and parameters of the nonlinear network can be determined.

Comparison with Linear and Computer Methods

Network analysis may be described as the process of relating the network currents to the source voltages and the element parameters. It is most directly accomplished by solving the circuit equations for the network currents and then mathematically analyzing the influence of each source voltage and element parameter on the network currents.

For linear networks, the usual phasor method provides an efficient tool for steady-state analysis when the source voltages can be expressed by the sum of a limited number of simple sinusoidal functions of time. Then, the general solution of the steady state is reduced to the solution of a correspondingly limited number of sets of simultaneous linear algebraic equations in terms of phasors with complex number coefficients.

For any type of network, after the time and effort necessary to establish the computer program has been expended, individual numerical solutions and families of such solutions can be rapidly calculated by automatic computers. Furthermore, unique solutions to specific problems will be obtained because the currents are traced, step by step, from initial to final conditions. However, network performance can then be analyzed only on the basis of results, rather than by means of functional forms developed in the process of solution.

For many nonlinear networks, the method described herein provides a process for the construction of steady-state solutions by phasor methods quite similar to those normally used in linear network analysis. In fact, if applied to a linear network, it reduces to the usual

Vol.
Pt. 1
1956

Paper 56-804, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 1, 1956; made available for printing May 10, 1956.

J. P. Neal is with the University of Illinois, Urbana, Ill.

This paper is based, in part, on a thesis presented to the Graduate College in partial fulfillment of the requirements for the Ph. D. degree in electrical engineering at the University of Illinois, 1955. The research was conducted under the supervision of Dr. G. H. Fett. Work on the thesis was supported, in part, by the General Electric Company.

phasor method of linear network analysis. The method has its limitations but does provide an engineering approach to and clearer understanding of many nonlinear problems.

Related Nonlinear Methods

The original approach to the solution of nonlinear differential equations, the method of perturbations, was described by Lindstedt and Poincare and has been extended by Van der Pol and others.[1–3] Kryloff and Bogoliuboff have more recently developed a method of equivalent linearization which shows how linearized elements can be developed to replace nonlinear network elements.[4,5] In effect, their linearized parameters are the coefficients of Fourier series representations of the network activities. Johnson, Thomsen and Schlesinger and others have shown how similar concepts can be successfully employed in the analysis of simple nonlinear networks.[6,7,8] These concepts are extended in this paper to multimesh-multinode networks, and simplified stability tests are developed and presented.

The steady state of any network is progressively complicated as nonlinear elements are introduced. The majority of the published work on nonlinear network analysis is devoted to networks containing only one nonlinear element. A phasor method of analysis of a multimesh-multinode network containing a single nonlinear element is very little more complicated than that of a single mesh or node-pair nonlinear network since, at any given frequency, the linear portion of any network can be replaced by a Thevenin or Norton equivalent. Such cases include only those nonlinear networks whose phasor-form equations for a single harmonic can be algebraically simplified to a single nonlinear phasor-form harmonic equation. More complicated networks can be solved by the method proposed herein.

General Description of the Method

Consider a nonlinear network operating in a specific steady state wherein each of the voltages and currents can be adequately represented by a finite series of simple sinusoidal functions of time. When these series representations of the mesh currents and source voltages have been substituted into the circuit equations for the nonlinear network, the terms can be simplified by means of trigonometric identities to simple sinusoidal time functions. This provides a set of equations consisting of sums of a finite number of simple sinusoidal time functions. These equations are now similar to the circuit equations of a linear network operating in a steady state and being driven at various frequencies simultaneously.

As may be done with any linear voltage equation, the terms of the same frequency from each of the simplified nonlinear equations can be separately equated. By this means, a set of single-frequency voltage equations can be derived for each harmonic occurring in the specific steady state and each derived set of voltage equations can be used as the defining equations of a corresponding harmonic network. Each term in each derived equation is interpreted as the voltage of a single linearized harmonic network element. In this manner, a set of harmonic networks are constructed which are equivalent, for that specific steady state, to the original nonlinear network.

The proposed phasor method of nonlinear network analysis follows the procedure just outlined except that the mesh currents and source voltages are represented in general form by finite trigonometric series having symbolic amplitudes and phase angles, and the specific harmonic frequencies found to be significant for the region of operation under investigation.

The selection of the significant harmonics is the crux of this method. It is based on previous experience, experiment, and preliminary analysis, with special attention to any resonant circuits in the harmonic networks. Very nonsinusoidal wave forms may often be represented by the sum of a very few harmonic components; if so, a few significant harmonics are often adequate for practical purposes.

In each simplified harmonic-circuit equation, the coefficient of each harmonic current is interpreted as the parameter of a linearized element in the corresponding harmonic network. At this point, since each simplified harmonic-circuit equation contains terms of a single frequency, the harmonic voltages, currents, and parameters may be represented by complex numbers and the balance of the analysis carried forward in phasor form.

Due to the nonlinearities of the original network, the parameters of the linearized harmonic elements can be functions of the amplitude and phase angle of any of the harmonic currents. Therefore evaluation of the amplitudes and phase angles of the harmonic currents requires the simultaneous solution of all of the circuit equations for each specific set of source voltages. This coupling which exists between the harmonic networks leads to multiple solutions for a given set of source voltages.

The uniqueness of a solution can be established by experiment or by tracing the transition from initial to final conditions by transient analysis, which is not within the scope of this paper. Stability of the various solutions is analyzed in the following paragraphs.

Laplace transforms, in particular the convolution integral, can be used to develop the form of the solutions even though nonlinear terms contain products or powers of currents.[9] However, that would be an unnecessary complication in this case where only the steady-state solutions are being considered and the significant harmonics are assumed initially.

STABILITY CONSIDERATIONS

For any specific solution to represent a steady state it must be stable. Instability may exist because random disturbances will not die with time or will cause a relatively gradual transition to a more stable state. Therefore, two types of stability tests are needed. The test which determines the instantaneous influence of small random disturbances is called the instant-stability test. The test which determines whether a relatively gradual transition to a more stable state will occur is called the average-stability test.

INSTANT-STABILITY TEST

The instant-stability test assumes that small random variations of each harmonic current encounter the same instantaneous influence as that encountered by that harmonic current in the linearized harmonic network. In other words, each harmonic network is examined for stability as though it were a linear network, each parameter being held constant at the value obtained for the specific solution under test.[5] For each set of harmonic-voltage equations in phasor form, p_h is substituted into the parameter determinant for $j\omega_h$ and the determinant is equated to zero and solved for the values of p_h. If the real component of any value of p_h is positive, the harmonic network is unstable for that specific solution and the solution does not represent a steady state. As with linear networks, instant stability can be verified directly from the determinantal equation in terms of p_h by application of the Routh-Hurwitz criterion.

AVERAGE-STABILITY TEST

The average-stability test assumes that, after instant stability has been verified, small random variations of the amplitudes and phase angles of the harmonic currents can be analyzed by the method of variations. The Poincare-Liapounoff criteria have been applied to the variational equations.[4] On that basis, the author developed certain tests as shown in the following. The derivation of the average-stability tests is outlined in Appendix I. The tests analyze the apparent input impedance facing each harmonic source, i.e., the complex number ratio of the harmonic source voltage to the harmonic input current. The average-stability conditions for each harmonic source and specific steady state are

$$\left[\frac{\partial(I^2R)}{\partial I}-\frac{\partial(IX)}{\partial\theta}\right]>0$$

$$\left\{\frac{\partial(IZ)^2}{\partial I}+2\left[\frac{\partial(IX)}{\partial I}\,\frac{\partial(IR)}{\partial\theta}-\frac{\partial(IR)}{\partial I}\,\frac{\partial(IX)}{\partial\theta}\right]\right\}>0$$

where

$Ee^{j\theta}$ = harmonic source voltage
$Ie^{j\theta}$ = harmonic input current
$Ze^{j\theta}$ = harmonic input impedance

and where R, X, and Z in the following equations are functions of I and θ

$$Z^2=R^2+X^2$$

$$P=I^2R$$

$$E=IZ$$

When the variations with respect to θ are negligible, the average-stability conditions reduce to

$$\frac{\partial P}{\partial I}>0 \text{ and } \frac{\partial E}{\partial I}>0$$

GRAPHIC EVIDENCE OF INSTABILITIES

At any harmonic source, zero or negative values of $\partial E/\partial I$, when linearized parameters are assumed constant, indicate instant instability.

If the variations with respect to θ are negligible, average stability is indicated at a particular harmonic source when the graphs of P and E versus I have positive slopes, providing instant stability has been verified.

Illustrative Application of the Method

A step-by-step illustrative application of this phasor method of nonlinear network analysis is presented in Appendix II. The network pertains to a practical problem which was proposed by the Ballast Department of the General Electric Company now located at Danville, Ill. That problem required the analysis of a peculiar instability which had caused undesirable oscillations of a frequency in the order of one cycle per second in an experimental fluorescent lamp and ballast system. The illustrative problem has previously been discussed.[11]

The analysis of the experimental system pertaining to the example shows that the illustrative network approached instant instability for higher source voltages because certain harmonic parameter-coefficients (r_{m2} and l_{m2}) became negative. Some discrepancy between measured and calculated results is possible because the nonlinear network of the example is not an exact mathematical model of the physical system to which it is said to be equivalent.

Conclusions

The method presented for analyzing multimesh-multinode nonlinear networks has a number of worth-while features. When the steady-state activities of a basic nonlinear network can be represented by the sum of a finite number of simple sinusoidal functions of time, the basic nonlinear network is suitable for this method. The prediction of the activities of that network can then be evaluated to any desired degree of accuracy if sufficient labor is expended.

The amount of labor required should not be overestimated. For repetitious calculations with similar networks the lengthy expansions need not be repeated. Identities relating polynomials of variables which are sums of simple sinusoidal functions of time to the expansions of those polynomials in terms of sums of simple sinusoidal functions of time, such those as listed in Appendix II, can be prepared or may be available. Then, the labor required for the evaluation of a specific steady-state solution may be comparable to that required for the evaluation of the steady-state solution of a linear network of similar plan and with the same number of harmonics present.

The crux of the method is the choice of the significant harmonics, since the reliability of the answers obtained is dependent on the correctness of the selection of the significant harmonics. As with many engineering problems, many false starts can be avoided if some concept of the form of the answer is known. The selection can be based on previous analyses of similar systems, preliminary analyses, or experiment.

Existence theorems for the solutions obtained by the method of equivalent linearization are given by Kryloff and Bogoliuboff in a series of papers, two of which have been translated and incorporated in a monograph by Lefschetz.[5] These theorems provide a basis for the method of approximate solutions which is a feature of the phasor method of nonlinear network analysis described herein.

The procedures used in this method are so similar to those used in linear network analysis that they may be followed without extensive additional training. The mathematics involved are primarily algebra, trigonometry, and the use of the phasor method of representing and manipulating simple sinusoidal functions of time.

The complete nonlinear influences are divided into component influences which can be successfully analyzed and evaluated. The linearized harmonic networks allow each harmonic to be analyzed somewhat independently as though the principle of superposition applied. Coupling between the linearized harmonic networks takes into account the conditions which prevented the use of the principle of superposition.

Appendix I. Development of Tests for Average Stability

The tests for the instant stability of the linearized harmonic networks assume that the linearized harmonic element parameters can be considered as constant over infinitesimal intervals. This is a reasonable assumption for instantaneous variations. However, the variations of the linearized parameters over finite intervals may not be negligible. The linearized parameter values are average functions of the harmonic current values, and therefore a test weighing the influence of changes in these parameter values is logically called an average-stability test.

The average-stability tests developed here apply separately to each linearized harmonic circuit and neglect any secondary influences due to coupling with other harmonic circuits or networks. That is, each harmonic circuit is considered as having linearized R, L, and C elements which are functions only of the current circulating throughout that harmonic circuit and influences from other harmonic circuits and networks are grouped together as a harmonic source voltage. On this basis, any harmonic circuit equation can be arranged as

$$L\ddot{q}+R\dot{q}+Sq=e$$

where q and e are periodic functions of time at the harmonic frequency ω, to which the circuit pertains. R, L, and S represent resistance, inductance, and elastance respectively, the linearized circuit parameters. The linearized circuit parameters may be functions of the amplitude Q and the phase θ of the harmonic charge q. Let

$$e=E\sin(\omega t+\theta)$$

$$q=Q\cos\omega t$$

$$\dot{q}=-\omega Q\sin\omega t$$

$$\ddot{q}=-\omega^2Q\cos\omega t$$

$$\theta=\tan^{-1}[\omega R/(\omega^2L-S)]$$

As a first step in this development, the inhomogeneous harmonic circuit equation will be rearranged in homogeneous form by transposing all terms to the left side of the equation. The equation is then in the form $L\ddot{q}+R_e\dot{q}+S_eq=0$, where

$$R_e=R+(E/\omega Q)\cos\theta$$

$$S_e=S-(E/Q)\sin\theta$$

The only case considered hereafter is where $R_e<4LS_e$. Then $q=Q\epsilon^{-\delta_p t}\cos\omega_p t$, where

$$\delta_p=(R_e/2L)$$

$$\omega_p=\sqrt{(S_e/L)-(R_e^2/4L_e^2)}$$

$$\theta=(\omega-\omega_p)t$$

For small variations of amplitude or phase, assume that

$$\dot{Q}=-\delta_pQ$$

$$\dot{\theta}=\omega-\omega_p$$

$$I=-\omega Q$$

$$\dot{I}=-\delta_pI$$

Thus

$$\dot{I}=-(R_e/2L)I=-(RI-E\cos\theta)/2L$$

$$\dot{\theta}=\omega-\{SL+(\omega LE/I)\sin\theta-[R-(E/I)\cos\theta]^2\}^{1/2}/L$$

Now, apply the method of variations as shown by Minorsky.[4] Let

$$I=I+\beta$$

$$\theta=\theta+\varphi$$

then

$$\dot{I}=\dot{\beta}$$

$$\dot{\theta}=\dot{\varphi}$$

The expansion of $\dot{I}$ and $\dot{\theta}$ in Taylor series about $I=I$ and $\theta=\theta$ gives

$$\dot{\beta}=\left(\dot{I}+\beta\frac{\partial\dot{I}}{\partial I}+\varphi\frac{\partial\dot{I}}{\partial\theta}+\ldots\right)$$

$$\dot{\varphi}=\left(\dot{\theta}+\beta\frac{\partial\dot{\theta}}{\partial I}+\varphi\frac{\partial\dot{\theta}}{\partial\theta}+\ldots\right)$$

Neglecting the implied higher-order terms and solving the equations simultaneously results in

$$\ddot{\varphi}-\left(\frac{\partial\dot{I}}{\partial I}+\frac{\partial\dot{\theta}}{\partial\theta}\right)\dot{\varphi}+\left(\frac{\partial\dot{I}}{\partial I}\frac{\partial\dot{\theta}}{\partial\theta}-\frac{\partial\dot{I}}{\partial\theta}\frac{\partial\dot{\theta}}{\partial I}\right)\varphi=\dot{I}\frac{\partial\dot{\theta}}{\partial I}$$

$$\ddot{\beta}-\left(\frac{\partial\dot{I}}{\partial I}+\frac{\partial\dot{\theta}}{\partial\theta}\right)\dot{\beta}+\left(\frac{\partial\dot{I}}{\partial I}\frac{\partial\dot{\theta}}{\partial\theta}-\frac{\partial\dot{I}}{\partial\theta}\frac{\partial\dot{\theta}}{\partial I}\right)\beta=\dot{\theta}\frac{\partial\dot{I}}{\partial\theta}$$

Considering the homogeneous form of the preceding equations, i.e., neglecting the right-hand terms, variations β and φ decrease with time and the periodic activity continues when

$$\left(\frac{\partial\dot{I}}{\partial I}+\frac{\partial\dot{\theta}}{\partial\theta}\right)<0$$

$$\left(\frac{\partial\dot{I}}{\partial I}\frac{\partial\dot{\theta}}{\partial\theta}-\frac{\partial\dot{I}}{\partial\theta}\frac{\partial\dot{\theta}}{\partial I}\right)>0$$

Partially differentiating $\dot{I}$ and $\dot{\theta}$ as necessary, and then setting

$$R_e=0$$

$$E\cos\theta=RI$$

$$E\sin\theta=\left(\omega L-\frac{S}{\omega}\right)I=XI$$

$$Z^2=R^2+X^2$$

gives

$$\frac{\partial\dot{I}}{\partial I}=-\frac{R}{2L}-\frac{I}{2L}\frac{\partial R}{\partial I}=-\frac{1}{2LI}\left[\frac{\partial(I^2R)}{\partial I}-IR\right]$$

$$\frac{\partial\dot{I}}{\partial\theta}=-\left(\frac{XI}{2L}+\frac{I}{2L}\frac{\partial R}{\partial\theta}\right)=-\frac{1}{2LI}\times\left[I^2X+I\frac{\partial(IR)}{\partial\theta}\right]$$

$$\frac{\partial\dot{\theta}}{\partial I}=\frac{1}{2LI}\left(X+I\frac{\partial X}{\partial I}\right)=-\frac{1}{2LI}-\left[\frac{\partial(IX)}{\partial I}\right]$$

$$\frac{\partial\dot{\theta}}{\partial\theta}=-\frac{1}{2L}\left(R-\frac{\partial X}{\partial\theta}\right)=-\frac{1}{2LI}\times\left[-\frac{\partial(IX)}{\partial\theta}+IR\right]$$

By substitution of these expressions and subsequent simplification, the average-stability conditions for a linearized harmonic network can be expressed as

$$\left[\frac{\partial(I^2R)}{\partial I}-\frac{\partial(IX)}{\partial\theta}\right]>0$$

$$\left\{\frac{\partial(IZ)^2}{\partial I}+2\left[\frac{\partial(IX)}{\partial I}\frac{\partial(IR)}{\partial\theta}-\frac{\partial(IR)}{\partial I}\frac{\partial(IX)}{\partial\theta}\right]\right\}>0$$

When R, X, and Z are practically independent of θ the average-stability conditions simplify to

$$\frac{dP}{\partial I}>0 \quad \text{where } P=I^2R$$

$$\frac{\partial E}{\partial I}>0 \quad \text{where } E=IZ$$

Appendix II. Illustrative Application of the Method

The phasor method of nonlinear network analysis is illustrated, step-by-step, by reviewing its application to a practical network equivalent to an experimental fluorescent lamp and ballast system.

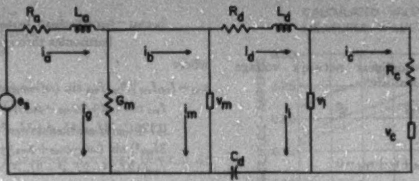


Fig. 1. Basic nonlinear example-network

Step 1. Specification of Basic Nonlinear Network

The plan of the basic nonlinear network of this example is illustrated in Fig. 1.

The element voltages of the basic nonlinear network are in volts

$$e_s=E_{s1}\sin(\omega t+\phi_{s1})$$

$$v_{Ra}=R_ai_a$$

$$v_{La}=L_a\dot{i}_a$$

$$v_{Rd}=R_di_d$$

$$v_{Ld}=L_d\dot{i}_d$$

$$v_{Gm}=R_mi_g$$

$$v_{Cd}=(1/C_d)q_d$$

$$v_{Rc}=R_ci_c$$

$$v_m=L_m\dot{i}_m$$

$$v_l=R_li_l+\sum_h A_{lh}\dot{i}_{lh}$$

$$v_c=(1/c)q_c$$

where

$\omega=377$ radians per second
$R_a=R_d=42$ ohms
$L_a=L_d=0.62$ henrys
$C_d=1.4\times10^{-6}$ farads
$G_m=(1/R_m)=65\times10^{-6}$ mhos
$R_c=1,250$ ohms
$c=(0.7\times10^{-6}/I_{l1})$ farads

$$L_m=(A_m+B_mi_m^2-C_mi_m^4)$$

where

$A_m=19$ henrys
$B_m=2\times10^4$ henrys per ampere^2
$C_m=15\times10^8$ henrys per ampere^4

$$R_l=A_l+B_li_l^2$$

where

$A_l=-(135/I_{l1})$ ohms
$B_l=(600/I_{l1}^3)$ volts per ampere^3
$A_{lh}=18,000/(h\omega I_{l1})^2$ henrys

The operating region of interest is $I_{a1}=0.22$ to 0.28 ampere. Subscripts relate to elements in Fig. 1. The positive reference of any v is at the tail of the associated current reference arrow, any e has the opposite reference.

The circuit currents of the basic nonlinear network are taken as i_g, i_m, i_l, and i_c. The other currents, in terms of the circuit currents, are

$$i_a=i_g+i_m+i_l+i_c$$

$$i_s = i_m + i_t + i_c$$

$$i_d = i_t + i_c$$

The basic nonlinear network voltage (circuit) equations are

$$v_{Rs} + v_{Ls} + v_m = e_s$$

$$-v_{Gm} + v_m = 0$$

$$-v_m + v_{Rd} + v_{Ld} + v_t + v_{Cd} = 0$$

$$-v_t + v_{Rc} + v_c = 0$$

Step 2. Selection of Significant Harmonics

Harmonic analyses and experiments with the system to which the example network is equivalent indicated four significant harmonics for the region of operation considered here: $h=0, 1, 2,$ and 3.

The general expression for each current is assumed as

$$i_k = I_{k0} + \sum_{h=1}^{3} i_{kh} = I_{k0} + \sum_{h=1}^{3} I_{kh} \sin \alpha_{kh}$$

$$= I_{k0} + \sum_{h=1}^{3} I_{kh} \sin (h\omega t + \phi_{kh})$$

where k identifies the associated element.

Furthermore, for any specific steady state

$$q_k = \sum_{h=0}^{3} q_{kh} = Q_{k0} + \sum_{h=1}^{3} Q_{kh} \cos \alpha_{kh}$$

$$= Q_{k0} + \sum_{h=1}^{3} (-I_{kh}/h\omega) \cos \alpha_{kh}$$

$$\dot{i}_k = \sum_{h=1}^{3} \dot{i}_{kh} = \sum_{h=1}^{3} h\omega I_{kh} \cos \alpha_{kh}$$

Step 3. Expansion of Basic Nonlinear Element Voltages

For each basic nonlinear element, the trigonometric representations of the currents are substituted into the element voltage function and the resultant expression is simplified to a series of simple sinusoidal functions of time by the use of trigonometric identities.

This step produces terms of frequencies additional to those previously classed as significant. If the significant harmonics were properly selected these additional terms may be discarded, as they are in this example. If there is doubt about the adequacy of the significant harmonics, it may be desirable to retain the additional frequency terms and observe their influence in subsequent steps.

The expanded, simplified, and subdivided nonlinear element voltages of the example nonlinear network for the region considered are shown in the following. To conserve space, the simplified trigonometric expansions are subdivided in anticipation of steps 4, 5, and 6. For each harmonic of each element voltage the terms are grouped into a harmonic source voltage and harmonic linearized element voltages as explained in step 4. The nonlinear element voltages v_m, v_t, and v_c expressed as sums of simple sinusoidal functions of time are, as follows

$$v_m = -e_{m1} + r_{m1}i_{m1} + l_{m1}\dot{i}_{m1} - e_{m2} + r_{m2}i_{m2} + l_{m2}\dot{i}_{m2} - e_{m3} + r_{m3}i_{m3} + l_{m3}\dot{i}_{m3} + \text{higher harmonics through the 15th,}$$

where

$$e_{m1} = I_{m2}I_{m3}\{B_m[I_{m0} \sin (\omega t - \phi_{m2} - \phi_{m3}) - I_{m2} \cos (\omega t + 2\phi_{m2} - \phi_{m3})] + 0.125C_m[4I_{m0}(4I_{m0}^2 + 3I_{m2}^2 + 3I_{m3}^2) \sin (\omega t - \phi_{m2} + \phi_{m3}) + I_{m2}(12I_{m0}^2 - 2I_{m2}^2 + 3I_{m3}^2) \times \cos (\omega t + 2\phi_{m2} - \phi_{m3})]\}$$

$$r_{m1} = \omega\{B_m[I_{m0}I_{m2} \cos (2\phi_{m1} - \phi_{m2}) - I_{m1}I_{m3} \sin (3\phi_{m1} - \phi_{m3})] + 0.0625C_m[8I_{m0}I_{m2}(4I_{m0}^2 + 5I_{m1}^2 + 3I_{m2}^2 + 6I_{m3}^2) \cos (2\phi_{m1} - \phi_{m2}) - 72I_{m0}I_{m1}I_{m2}I_{m3} \cos (\phi_{m1} + \phi_{m2} - \phi_{m3}) + 2I_{m1}^2I_{m2}^2 \sin 2(2\phi_{m1} - \phi_{m2}) + 3I_{m1}I_{m3}(8I_{m0}^2 + I_{m1}^2 - 4I_{m2}^2 - 2I_{m3}^2) \sin (3\phi_{m1} - \phi_{m3}) + 4I_{m1}I_{m2}^2I_{m3} \sin (\phi_{m1} - 2\phi_{m2} + \phi_{m3}) + 6I_{m2}^2I_{m3}^2 \sin 2(\phi_{m1} + \phi_{m2} - \phi_{m3})]\}$$

$$l_{m1} = A_m + 0.25B_m[4I_{m0}^2 + I_{m1}^2 + 2I_{m2}^2 + 3I_{m3}^2 - I_{m0}I_{m2} \sin (2\phi_{m1} - \phi_{m2}) + I_{m1}I_{m3} \cos (3\phi_{m1} - \phi_{m3})] - 0.0625 \times C_m[16I_{m0}^4 + 24I_{m0}^2I_{m1}^2 + 48I_{m0}^2I_{m2}^2 + 2I_{m1}^4 + 12I_{m1}^2I_{m2}^2 + 12I_{m1}^2I_{m3}^2 + 6I_{m2}^4 + 24I_{m2}^2I_{m3}^2 + 6I_{m3}^4 + 8I_{m0}I_{m2}(4I_{m0}^2 + 7I_{m1}^2 + 3I_{m2}^2 + 6I_{m3}^2) \sin (2\phi_{m1} - \phi_{m2}) - 24I_{m0}I_{m1}I_{m2}I_{m3} \times \sin (\phi_{m1} + \phi_{m2} - \phi_{m3}) - 2I_{m1}^2I_{m2}^2 \times \cos 2(2\phi_{m1} - \phi_{m2}) + 28I_{m1}I_{m2}^2I_{m3} \times \cos (\phi_{m1} - 2\phi_{m2} + \phi_{m3}) - I_{m1}I_{m3} \times (24I_{m0}^2 + 5I_{m1}^2 - 12I_{m2}^2 - 6I_{m3}^2) \times \cos (3\phi_{m1} - \phi_{m3}) - 6I_{m2}^2I_{m3}^2 \times \cos 2(\phi_{m1} + \phi_{m2} - \phi_{m3})]$$

$$e_{m2} = I_{m0}I_{m1}\{2B_mI_{m3} \sin (2\omega t - \phi_{m1} + \phi_{m3}) + C_m[I_{m1}(2I_{m0}^2 + I_{m1}^2 + 3I_{m3}^2) \times \sin 2(\omega t + \phi_{m1}) - I_{m3}(4I_{m0}^2 + 3I_{m1}^2 + 3I_{m3}^2) \sin (2\omega t - \phi_{m1} + \phi_{m3})]\}$$

$$r_{m2} = \omega\{-B_mI_{m1}I_{m3} \sin (\phi_{m1} - 2\phi_{m2} + \phi_{m3}) + 0.125C_m[-12I_{m0}I_{m2}(I_{m1}^2 + 2I_{m3}^2) \cos (2\phi_{m1} - \phi_{m2}) + 24I_{m0}I_{m1} \times I_{m2}I_{m3} \cos (\phi_{m1} + \phi_{m2} - \phi_{m3}) - I_{m1}^4 \sin 2(2\phi_{m1} - \phi_{m2}) + 4I_{m1}^3I_{m3} \sin (3\phi_{m1} - \phi_{m3}) + 6I_{m1}^2I_{m3}^2 \sin 2(\phi_{m1} + \phi_{m2} - \phi_{m3}) + 4I_{m1}I_{m3}(12I_{m0}^2 + 3I_{m1}^2 + 2I_{m2}^2 + 3I_{m3}^2) \sin (\phi_{m1} - 2\phi_{m2} + \phi_{m3})]\}$$

$$l_{m2} = A_m + 0.25B_m[4I_{m0}^2 + 2I_{m1}^2 + I_{m2}^2 + 2I_{m3}^2 + 2I_{m1}I_{m3} \cos (\phi_{m1} - 2\phi_{m2} + \phi_{m3})] - 0.0625C_m[16I_{m0}^4 + 48I_{m0}^2I_{m1}^2 + 24I_{m0}^2I_{m2}^2 + 48I_{m0}^2I_{m3}^2 + 6I_{m1}^4 + 12I_{m1}^2I_{m2}^2 + 24I_{m1}^2I_{m3}^2 + 2I_{m2}^4 + 12I_{m2}^2I_{m3}^2 + 6I_{m3}^4 + 12I_{m0}I_{m2}(I_{m1}^2 + 2I_{m1}I_{m3} - 2I_{m3}^2) \sin (2\phi_{m1} - \phi_{m2}) + 72I_{m0}I_{m1}I_{m2}I_{m3} \sin (\phi_{m1} + \phi_{m2} - \phi_{m3}) - I_{m1}^4 \cos 2(2\phi_{m1} - \phi_{m2}) - 12I_{m1}^3I_{m3} \cos (3\phi_{m1} - \phi_{m3}) - 6I_{m1}^2I_{m3}^2 \cos 2(\phi_{m1} + \phi_{m2} - \phi_{m3}) + 4I_{m1}I_{m3}(12I_{m0}^2 + 3I_{m1}^2 + 4I_{m2}^2 + 3I_{m3}^2) \cos (\phi_{m1} - 2\phi_{m2} + \phi_{m3})]$$

$$e_{m3} = \omega I_{m1}\{0.25B_m[-12I_{m0}I_{m2} \sin (3\omega t + \phi_{m1} + \phi_{m2}) + I_{m1}^2 \cos 3(\omega t + \phi_{m1}) - 3I_{m2}^2 \cos (3\omega t - \phi_{m1} + 2\phi_{m2})] - 0.1825C_m[-8I_{m0}I_{m2}(4I_{m0}^2 + 3I_{m1}^2 + 3I_{m2}^2) \sin (3\omega t + \phi_{m1} + \phi_{m2}) + I_{m1}^2(8I_{m0}^2 + I_{m1}^2 + 4I_{m2}^2) \times \cos 3(\omega t + \phi_{m1}) - 2I_{m2}^2(4I_{m0}^2 + 3I_{m1}^2 + 3I_{m2}^2) \cos (3\omega t - \phi_{m1} + 2\phi_{m2})]\}$$

$$r_{m3} = 0.375\omega C_m[24I_{m0}I_{m1}^2I_{m2} \cos (2\phi_{m1} - \phi_{m2}) - 12I_{m0}I_{m1}I_{m2}I_{m3} \cos (\phi_{m1} + \phi_{m2} - \phi_{m3}) - 4I_{m0}I_{m2}^3 \cos (3\phi_{m2} - \phi_{m3}) - I_{m1}^3I_{m3} \sin (3\phi_{m1} - \phi_{m3}) - 3I_{m1}^2I_{m2}^2 \sin 2(\phi_{m1} + \phi_{m2} - \phi_{m3}) - 3I_{m1}I_{m2}^2I_{m3} \sin (\phi_{m1} - 2\phi_{m2} + \phi_{m3})]$$

$$l_{m3} = A_m + 0.25B_m(4I_{m0}^2 + 2I_{m1}^2 + 2I_{m2}^2 + I_{m3}^2) - 0.125C_m[4I_{m0}^2(2I_{m0}^2 + 6I_{m1}^2 + 6I_{m2}^2 + 3I_{m3}^2) + 3I_{m1}^2(I_{m1}^2 + 4I_{m2}^2 + 2I_{m3}^2) + I_{m2}^2(I_{m2}^2 + 2I_{m3}^2) + I_{m3}^4 + 48I_{m0}I_{m1}I_{m2}I_{m3} \sin (\phi_{m1} + \phi_{m2} - \phi_{m3}) + 4I_{m0}I_{m2}^3 \sin (3\phi_{m2} - 2\phi_{m3}) - 3I_{m1}^3I_{m3} \cos (3\phi_{m1} - \phi_{m3}) - 3I_{m1}^2I_{m2}^2 \cos 2(\phi_{m1} + \phi_{m2} - \phi_{m3}) - 12I_{m1}I_{m2}^2I_{m3} \cos (\phi_{m1} - 2\phi_{m2} + \phi_{m3})]$$

$$v_t = -e_{t0} + r_{t0}i_{t0} - e_{t1} + r_{t1}i_{t1} + l_{t1}\dot{i}_{t1} - e_{t2} + r_{t2}i_{t2} + l_{t2}\dot{i}_{t2} - e_{t3} + r_{t3}i_{t3} + l_{t3}\dot{i}_{t3} + \text{higher harmonics through the ninth harmonic.}$$

where

$$e_{t0} = -0.75B_tI_{t1}I_{t2}[I_{t1} \sin (2\phi_{t1} - \phi_{t2}) + I_{t3} \times \sin (\phi_{t1} + \phi_{t2} - \phi_{t3})] = E_{t0}$$

$$r_{t0} = A_t + B_t(I_{t0}^2 + 1.5I_{t1}^2 + 1.5I_{t2}^2 + 1.5I_{t3}^2)$$

$$e_{t1} = -0.75B_tI_{t2}I_{t3}[4I_{t0} \cos (\omega t - \phi_{t2} + \phi_{t3}) + I_{t2} \sin (\omega t + 2\phi_{t2} - \phi_{t3})]$$

$$r_{t1} = A_t + 0.75B_t[4I_{t0}^2 + I_{t1}^2 + 2I_{t2}^2 + 2I_{t3}^2 + 4I_{t0}I_{t2} \sin (2\phi_{t1} - \phi_{t2}) - I_{t1}I_{t3} \cos (3\phi_{t1} - \phi_{t3})]$$

$$l_{t1} = A_{t1} + (3/4\omega)B_t[4I_{t0}I_{t2} \cos (2\phi_{t1} - \phi_{t2}) + I_{t1}I_{t3} \sin (3\phi_{t1} - \phi_{t3})]$$

$$e_{t2} = 1.5B_tI_{t0}I_{t1}[I_{t1} \cos 2(\omega t + \phi_{t1}) - 2I_{t3} \times \cos (2\omega t - \phi_{t1} + \phi_{t3})]$$

$$r_{t2} = A_t + 0.75B_t[4I_{t0}^2 + 2I_{t1}^2 + I_{t2}^2 + 2I_{t3}^2 + 2I_{t1}I_{t3} \cos (\phi_{t1} - 2\phi_{t2} + \phi_{t3})]$$

$$l_{t2} = A_{t2} + (3/4\omega)B_tI_{t1}I_{t3} \sin (\phi_{t1} - 2\phi_{t2} + \phi_{t3})$$

$$e_{t3} = 0.25B_tI_{t1}[12I_{t0}I_{t2} \cos (3\omega t + \phi_{t1} + \phi_{t2}) + I_{t1}^2 \sin 3(\omega t + \phi_{t1}) - 3I_{t2}^2 \sin (3\omega t - \phi_{t1} + 2\phi_{t2})]$$

$$r_{t3} = A_t + 0.75B_t(4I_{t0}^2 + 2I_{t1}^2 + 2I_{t2}^2 + I_{t3}^2)$$

$$l_{t3} = A_{t3}$$

$$v_c = (1/c)(q_{c0} + q_{c1} + q_{c2} + q_{c3})$$

Step 4. Specification of Harmonic

Each harmonic network has the same plan and element locations as the basic nonlinear network. The voltages of the linear harmonic elements are the same as those in the basic nonlinear network. The voltages of the linearized harmonic elements are the harmonic terms of the step 3 expansions for the corresponding basic nonlinear elements.

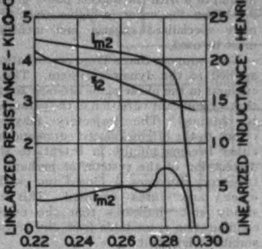


Fig. 2. Calculated and measured input currents and powers versus input voltage for the example network

Every basic nonlinear element is represented in a harmonic network by a series combination including a linearized harmonic source and linearized passive elements, depending upon the terms forming the element voltages. Terms which are not functions of the instantaneous harmonic current designate a linearized harmonic source. Terms which are functions of the instantaneous harmonic current designate a linearized resistance, inductance, or capacitance, according to the form of the terms.

For example, for the first harmonic network, v_{m1} includes all the first harmonic terms of v_m. Of the terms included in v_{m1}, those which include i_{m1}, $\dot{i}_{m1}$, or q_{m1} explicitly constitute linearized passive voltages, the other terms constitute a harmonic source voltage which is designated e_{m1}. The coefficients of i_{m1} constitute a linearized resistance. The coefficients of $\dot{i}_{m1}$ constitute a linearized inductance. Because q_{m1} does not appear in v_{m1}, there is no linearized capacitance in this instance.

Step 5. Development of Harmonic Phasors

As explained and demonstrated in the preceding steps, the voltage of each harmonic network element consists of a harmonic source voltage or of a linearized coefficient or parameter multiplying a current. Therefore, as with linear networks, each harmonic voltage and current can be represented, for a specific steady state, as a stationary phasor in complex or polar form for the particular harmonic frequency.

The phasor representations for the voltages and currents of the example, with the use of maximum-value phasors and the parameters listed in step 3, are as follows.

THE 0TH (D-C) HARMONIC NETWORK

$\omega_0=0$
$E_{s0}=0$
$V_{Ra0}=R_aI_{a0}$
$V_{La0}=0$
$V_{Rd0}=R_dI_{d0}$
$V_{Ld0}=0$
$V_{Gm0}=R_mI_{g0}$
$V_{Cd0}=(1/C_d)Q_{d0}$
$V_{Rc0}=R_cI_{c0}$
$V_{m0}=0$
$V_{t0}=-E_{t0}+r_{t0}I_{t0}$
$V_{c0}=(1/c)Q_{c0}$

where the representation of i_{k0} is I_{k0}.

THE HTH HARMONIC NETWORK ($h=1, 2, 3$)

$\omega_h=h\omega$
$\bar{E}_{sh}=E_{sh}e^{j\phi_{sh}}$
$\bar{V}_{Rah}=R_a\bar{I}_{ah}$
$\bar{V}_{Lah}=j\omega_hL_a\bar{I}_{ah}$
$\bar{V}_{Rdh}=R_d\bar{I}_{dh}$
$\bar{V}_{Ldh}=j\omega_hL_d\bar{I}_{dh}$
$\bar{V}_{Gmh}=R_m\bar{I}_{gh}$
$\bar{V}_{Cdh}=(1/j\omega_hC_d)\bar{I}_{dh}$
$\bar{V}_{Rch}=R_c\bar{I}_{ch}$
$\bar{V}_{mh}=-\bar{E}_{mh}+r_{mh}\bar{I}_{mh}+j\omega_hl_{mh}\bar{I}_{mh}$
$\bar{V}_{th}=-\bar{E}_{th}+r_{th}\bar{I}_{th}+j\omega_hl_{th}\bar{I}_{th}$
$\bar{V}_{ch}=(1/j\omega_hc)\bar{I}_{ch}$

where the phasor representation of i_{kh} is $\bar{I}_{kh}=I_{kh}e^{\phi_{ikh}}$.

Step 6. Development of Harmonic Voltage (Circuit) Equations

After the harmonic networks have been completely specified, the values of the harmonic source voltages and the linearized harmonic element parameters have been established, and the harmonic voltages represented in phasor form, the harmonic voltage (circuit) equations are written in phasor form in the same manner as for linear networks. There are H sets of voltage equations corresponding to the H harmonic networks.

Coupling between the networks is evidenced by the appearance of the harmonic current amplitudes and phase angles of other harmonic networks in the expressions for the source voltages and linearized parameters of any given harmonic network. It is because of this interharmonic coupling that the principle of superposition does not apply to nonlinear networks.

The voltage (circuit) equations for the example, in general form for the hth harmonic network with p written for $j\omega_h$, are

$$\begin{bmatrix} R_a+R_m+pL_a & R_a+pL_a & R_a+pL_a & R_a+pL_a \\ -R_m & r_{mh}+pl_{mh} & 0 & 0 \\ 0 & -r_{mh}-pl_{mh} & R_d+r_{th}+p(L_d+l_{th})+(1/pC_d) & R_d+pL_d+(1/pC_d) \\ 0 & 0 & -r_{th}-pl_{th} & R_c+(1/pc) \end{bmatrix}\times \begin{bmatrix} I_{gh} \\ I_{mh} \\ I_{th} \\ I_{ch} \end{bmatrix}=\begin{bmatrix} E_{sh} \\ E_{mh} \\ E_{th}-E_{mh} \\ -E_{th} \end{bmatrix}$$

Step 7. Evaluation of Unknown Currents

Because of the interharmonic coupling, it is necessary to solve all H sets of the harmonic voltage equations simultaneously.

In the example, the range of currents within the operating region concerned was known approximately from previous experimental work, and manual solutions were readily, though tediously, calculated by assuming values for I_{11}; minor terms were ignored in the linearized expressions and the corresponding approximate values of the other currents were calculated. A graph of the values of I_{a1}, I_{a3}, and P_{s1} versus E_{s1} is shown in Fig. 2. Values of I_a and P_s which were measured on the physical system to which the example network is supposed to be equivalent are also shown versus E_s in Fig. 2, for comparison. During stable operation the 0th and 2nd harmonic currents are zero.

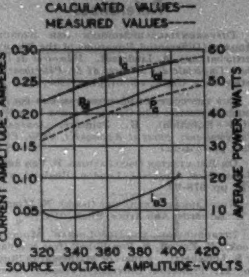


Fig. 3. Linearized parameters of the quiescent second-harmonic example-network versus the first-harmonic input current

Step 8. Verification of Stability

To confirm the fact that a specific solution represents a specific steady state, it is necessary to verify the instant and average stabilities of every harmonic network for the particular specific solution by means of the tests described in the text.

For the example network, instant and average stability were verified for every significant harmonic network within the region $0.22I<_{a1}<0.28$ ampere. Within that region every parameter of every significant harmonic network had real positive values, verifying instant stability. The verifications of average stability are illustrated in Fig. 2, where the input current and power are plotted versus source voltage. Within the region mentioned the partial derivatives of input power and source voltage with respect to input current are positive and finite for the first harmonic. The third harmonic input current is also plotted versus the input source voltage; the final upturn of that curve suggests the approach to instant instability which occurs thereafter in the second harmonic network.

The occurrence of instant stability in the second harmonic network when I_{a1} is greater than 0.29 ampere is indicated in Fig. 3. Then, l_{m2} and r_{m2} become negative if the zero and second harmonic networks are quiescent.

References

1. Differentialgleichungen der Storungstheorie (Differential Equations of the Theory of Perturbations), **A. Lindstedt.** *Mémoires de l'Académie Impériale des Sciences de St. Pétersbourg,* St. Petersbourg, USSR, vol. 31, pt. 4, 1883, pp. 1–20.

2. Sur les courbes définies par les équation différentielles (On Curves Defined by a Differential Equation), **H. Poincaré.** *Journal de Mathématiques pures et Appliquées,* Paris, France, 1886, pp. 151–217.

3. On Relaxation Oscillations, **B. Van der Pol.** *Philosophical Magazine,* London, England, vol. 2, 1926, pp. 978–92.

4. Non-Linear Mechanics (book), **N. Minorsky.** J. W. Edwards, Ann Arbor, Mich., 1947.

5. Introduction to Non-Linear Mechanics (book), **N. Kryloff, N. Bogoliuboff** (free translation by **S. Lefschetz**). Annals of Mathematics Studies, No. 11, Princeton University Press, Princeton, N. J., 1947.

6. Sinusoidal Analysis of Feedback-Control Systems Containing Nonlinear Elements, **E. Calvin Johnson.** *AIEE Transactions,* vol. 71, pt. II, 1942, pp. 169–81.

7. Graphical Analysis of Nonlinear Circuits Using Impedance Concepts, **J. S. Thomsen.** *Journal of Applied Physics,* New York, N. Y., vol. 24, 1953, p. 1379.

8. Analysis of Nonlinear Circuits Using Impedance Concepts, **J. S. Thomsen, S. P. Schlesinger.** *Transactions,* Professional Group on Circuit Theory, Institute of Radio Engineers, New York, N. Y., vol. CT-2, 1955, pp. 271–78.

9. Modern Operational Mathematics in Engineering (book), **R. V. Churchill.** McGraw-Hill Book Company, Inc., New York, N. Y., 1944, pp. 36–38.

10. Feedback Control Systems (book), **G. H. Fett.** Prentice-Hall, Inc., New York, N. Y., 1955, pp. 105–10.

11. The Phasor Method of Nonlinear Network Analysis, **J. P. Neal.** *Ph.D. Thesis,* University of Illinois, Urbana, Ill., 1955.

Discussion

G. H. Fett (University of Illinois, Urbana, Ill.): Dr. Neal has developed a compact solution to the nonlinear circuit problem which follows the technique of phasors familiar to all electrical engineers. The instant- and average-stability tests given provide a direct method of determining whether or not a particular solution is stable and physically realizable. They take the place of considerations of stability which can be obtained from the phase plane analysis for a system describable in terms of a single co-ordinate and its time derivative.

A comparison between the two methods may help to point out the advantage of the method given by Dr. Neal. In the phase plane the steady-state solution, which is the one obtained by Dr. Neal, is evidenced by a closed trajectory. Suppose that such a trajectory has been drawn for the system described by the harmonic networks, and that a small perturbation is suddenly interjected. Since the system is assumed to have a steady state, and the perturbation is small, the succeeding trajectories will approach the original one after a period of time. The parameters of the system are assumed to have changed very little and, indeed, for these average values of the parameters the Routh or Hurwitz criteria are satisfied for stability. This represents the instant stability of the paper. In connection with this statement it should be pointed out that the second sentence in the second paragraph under step 8 in Appendix II is pertinent when the equation is a polynomial in p with the highest power given as 2. When the order is higher than 2 the more specialized criteria just mentioned must be used.

Now, suppose that a large perturbation is applied to the dynamic system. The excursions in the phase plane are now so large that a significant variation in the parameters is obtained. The trajectory may not return to the original closed curve and may either become infinite in extent, to show destruction of the system, or perhaps the trajectory becomes a closed one with a much larger area enclosed. This new steady state indicates that the original configuration is unstable from the average-stability viewpoint.

To make a determination of instant and average stability, with the use of the phase plane diagram, required that the entire range of values of parameters be investigated. Furthermore, if the system had been one which required more than two coordinates, or especially more than three, the visualization of the phase space picture would be difficult. The method of Dr. Neal is straightforward, even for these more difficult problems.

It should be pointed out that Dr. Neal has taken a lot of the tedium out of the solution to the nonlinear problem with the material which appears in step 3 of Appendix II. The trigonometric terms which must be expanded and collected are already arranged for the user, including the appropriate phase angles.

Influence of Noise on Telephone Signaling Circuit Performance

LAURANCE A. WEBER
ASSOCIATE MEMBER AIEE

Synopsis: This paper reports results of a comprehensive study of influence of noise on signaling performance of the major a-c telephone signaling systems in use by the Bell System. A problem exists in the evaluation of signaling circuit performance with regard to characterization of both noise and pulsing accuracy. It appears that a good way to characterize noise is to use the 1% (per cent) point on a cumulative level distribution taken with a level distribution recorder with a 10-ms (millisecond) time constant. Pulsing performance can best be characterized by obtaining a distribution of pulse distortion while pulsing over a noise-disturbed circuit. The 2,600-cycle in-band signaling system performs satisfactorily for signal-to-1%-10-ms-noise ratios of about 0 db (decibels). The *N* carrier out-of-band system is about 6 db better than the 2,600-cycle system. While comparison is difficult due to different circuit action, the 1,600-cycle in-band system is about 10 db more sensitive to noise than the 2,600-cycle system. Some modification of signal-to-guard ratios in the 2,600-cycle system appears to result in performance which is 6 to 8 db better. Tests showed that compandor action is not detrimental to in-band signaling performance.

IN THE Bell System, signaling over toll telephone trunks is now largely obtained by a-c methods. The power of a-c signals usually is about the same order of magnitude as the speech power over the trunk so that it is true for signaling as well as speech transmission that influence of noise on performance is a limiting factor in system design. Optimum design of a toll trunk requires that the speech and signaling system withstand about the same noise before failure. Increased use is made of carrier circuits to derive toll channels. On these carrier circuits the length of repeater section is determined by the difference between maximum transmitted level and minimum received level. The first of these is a function of the power-handling capacity of the carrier system and by interference into paralleling systems. The second is a function of the system susceptibility to noise. This makes the adequate specification of performance in the face of noise of signaling circuits increasingly important. This paper reports results of a comprehensive study of the influence of noise on signaling performance of the major a-c signaling systems used by the

---

Paper 56-784, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 26, 1956; made available for printing April 30, 1956.

Laurance A. Weber is with Bell Telephone Laboratories, Inc., New York, N. Y.

The author acknowledges the contributions of W. D. Garland and J. F. Griffin, Jr., Bell Telephone Laboratories, to the results reported herein; A. J. Aikens provided the tape recordings of noise used in the tests.

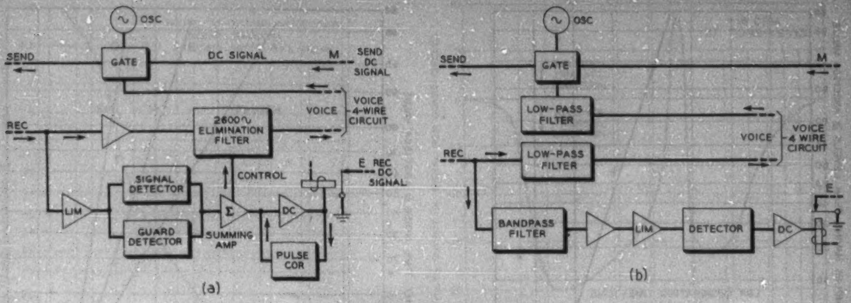


Fig. 1. Block diagram of signaling circuits

a—In-band type b—Out-of-band type

Bell System. These are the 1,600-cycle and the 2,600-cycle single-frequency systems;[1] and the *N*- and *O*-carrier 3,700-cycle single-frequency system.[2]

Toll-telephone signaling schemes in the United States have generally used simple amplitude-modulation subcarrier methods. The circuits in this study are of this type. Some important use of frequency-modulation and frequency-shift methods have been reported.[3,4] It has been shown that such circuits have a slight noise advantage over simple amplitude-modulation methods. Generally, however, the amplitude-modulation systems are more attractive from a cost standpoint and so have so far been used in all major trunk signaling schemes in the Bell System. Future developments may, of course, change this picture. The amplitude-modulation signaling schemes in use can be divided into in-speech-band circuits and out-of-speech-band circuits.

In-band circuits signal on a binary basis with a single frequency in the voice band (usually 1,600 or 2,600 cycles). This tone is removed during the normal subscriber talking interval so that no interference is obtained and the full voice-frequency range is available for talking. Where certain calls, e.g., to intercepting operators, require conversation while signaling tone is transmitted, the voice-frequency band is modified to provide elimination of the signaling frequency toward the customer. In this type of design it is necessary to prevent signal imitation by speech from falsely operating the signaling receiver. Special problems also arise when sending line-busy flashes and audible busy back tone. The outstanding advantage of the scheme is that it can be applied over multilink connections of differing facilities such that d-c signal repetition is not required at intermediate points. For example, a trunk from Boston, Mass., to Chicago, Ill., may consist of a *K*-carrier channel from Boston to New York, N. Y., and a *TD2* microwave channel from New York to Chicago with voice-frequency interconnection of the two channels on a permanent basis at New York. For this case, in-band signaling allows the signaling channel to be interconnected with the voice channel at New York without additional equipment. This produces a degree of flexibility, and also reduces the signaling equipment when the trunk circuit contains more than one link. The 1,600- and 2,600-cycle single-frequency systems are of the in-band type.

Out-of-band schemes are characterized by a division of the available transmission band width into a voice channel and a signaling channel. The 3,700-cycle single-frequency system used with *N*- and *O*-carrier systems is typical of this type.[5] In this system the voice band is limited to about 3,100 cycles by means of low-pass filters and a channel at 3,700 cycles is used for on–off binary signaling. This general scheme is attractive where the signaling circuit is integral with the carrier system physically since low cost and simplicity usually result. These carrier systems generally are short-haul types where it is not expected that many multilink facilities are required. Fig. 1 shows a block diagram of the general arrangement of the in-band and out-of-band signaling circuits.

Characterization Problems

In connection with studies of new applications of carrier methods, a comprehensive series of tests was performed on the three signaling circuits mentioned to characterize their performance against noise of the type usually encountered in telephone practice. This performance characterization involves three problems:

1. Specification of noise in measurable terms related to signaling performance.
2. Specification of signaling performance.
3. Conducting tests to relate problems 1 and 2.

Measuring Noise

It would be desirable to place quantitative "tag" on a given sample of noise such that pulsing performance of a signaling circuit in the face of this noise will be identical with the performance of the circuit with other samples of noise having this same quantitative tag. Normal telephone noise measurement in db with respect to reference noise (dbRN) or db adjusted by a *2B* noise measuring set is not adequate because the set has an integrating time of about 200 ms to simulate the response of the ear. Usually the response of a signaling circuit is of the order of 15 to 40 ms. An rms measurement is not adequate because noise often is very peaked in character with low average levels between peaks. An rms value would give too low a measure of the noise as it affects signaling. The approach used here is to obtain the statistics of the noise sample by sampling quasipeak levels over a time and empirically selecting a significant point on this distribution as the suitable measure of the noise. This point is chosen by running a number of samples of noise with differing level distributions into a signaling circuit, obtaining equal pulsing performance for each sample by adjusting the noise level with attenuators, then finding a common level value on the distribution curve. If this value of noise level (say, in dbRN) is used to characterize the noise, then differing noise sam-

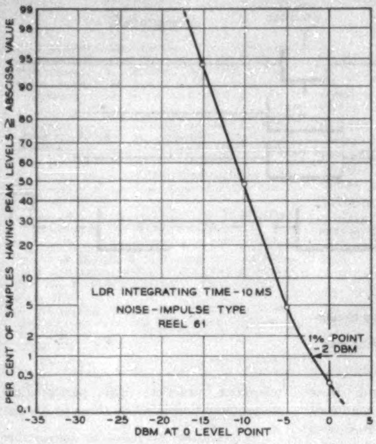


Fig. 2. Typical noise distribution from LDR readings

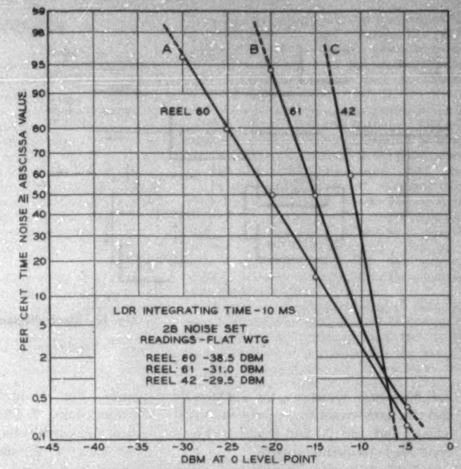


Fig. 3. LDR measurements of different noise samples; levels adjusted for equal 2,600-cyle signaling performance

ples having the same "level" will give essentially equal signaling performance. It is hoped, of course, that this "equal performance" specification for noise will be general enough to apply to all signaling circuits tested. In this study, it seemed to be the case.

A device which is convenient to use to obtain statistics of noise samples is a level distribution recorder (LDR) which has been used in previous studies.[6] This instrument samples a signal once each second, determining in this interval the peak level present over integrating times of 1, 10, or 200 ms. A bank of registers calibrated in 5-db steps indicates directly the statistical distribution of the energy level of the source. Fig. 2 shows a distribution obtained by this means of noise on a typical circuit. The plot is made on arithmetic probability paper on which linearity indicates a normal distribution.

Some distribution curves of different noise samples with levels adjusted for equal signaling performance are shown in Fig. 3. Here the differing character of the noise is readily apparent. Curve *A* represents noise having a low background level; curve *C* represents noise with a relatively continuous high level. In each case here, however, it appears that as far as signaling performance is concerned each of these samples may be characterized by the level corresponding to the 1% distribution point on the LDR curve with a 10-ms integration time. Hence noise levels can be quoted as dbRN (0 dbRN = 90 db below 1 milliwatt) at the 1% point on a 10-ms LDR distribution curve for the particular noise. To indicate the inadequacy of the *2B* noise measuring set readings, Fig. 3 also shows points corresponding to the readings on this set for the three noise samples. As can be seen, the readings differ by as much as 15 db and yet signaling is affected about the same for each noise sample.

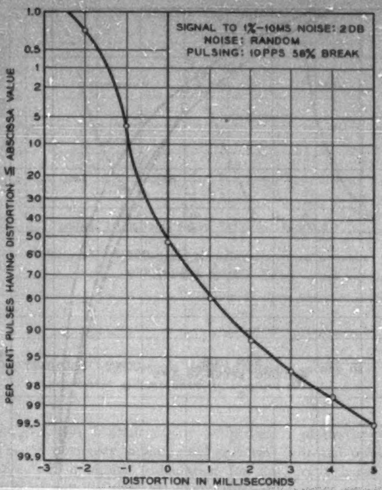
Noise is somewhat elusive on telephone lines so that obtaining samples for testing purposes is a problem. This study utilized tape-recorded samples which were made at appropriate periods during high noise intensity on the lines. LDR distributions obtained while recording and from the playback of the tape served to assure faithful recording. Two types of noise are of interest in the signaling problem: noise originating from atmospheric static and noise generated by switching transients in central offices. The static noise generally has larger peak amplitudes but central office noise is bothersome over longer intervals during the year. Both types of noise must be considered in evaluating signaling circuit performance.

Measuring Signaling Performance

The condition most influenced by noise for a signaling circuit is the dial pulsing condition. Here rapid alternations between the binary on–off conditions transmit the digital information of a telephone user's called number; an error will cause misdirection of the call. It is natural then for the performance of the circuit to be evaluated under pulsing conditions. In the absence of noise a signaling circuit will pulse at a speed and on–off ratio determined by the source of pulses feeding it. The effect of noise will be a jitter in the on–off ratio of varying amounts. This jitter is measured in terms of per-cent break distortion or bias, where, in the system studied here, per-cent break is defined as

$$\text{per cent break} = \frac{\text{tone-on time}}{\text{tone-on time} + \text{tone-off time}}$$

For high noise levels the single pulse distortion will be high and more pulses will be distorted. At low noise levels the reverse will be true. A cumulative distribution plot of numbers of pulses distorted by given amounts on probability paper is a convenient way to characterize signaling circuit performance. Families of such curves can be plotted, showing distortion characteristics for different noise levels and for differing noise samples. Information of this type is a valuable tool to use in maximizing certain circuit parameters to obtain optimum noise immunity. Useful comparisons can be made between signaling circuits by adjusting noise levels so that equal performance plots are obtained and a rough equivalency is established between circuits. This is the pro-



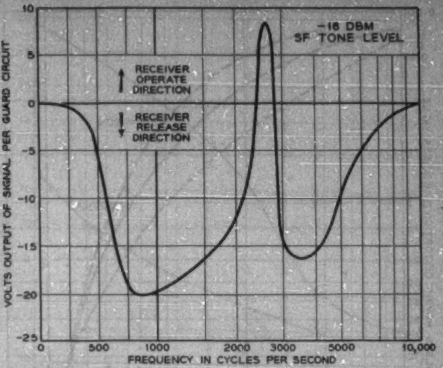


Fig. 4 (left). N-carrier 3,700-cycle signaling circuit pulsing distortion distribution

Fig. 7 (above right). Signal-guard characteristics of 2,600-cycle signaling circuit

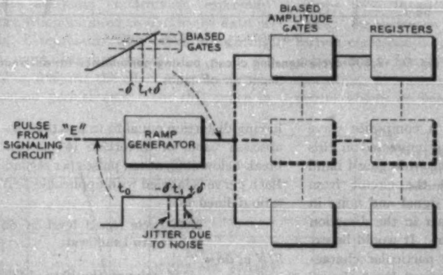

BIASED GATES
BIASED AMPLITUDE GATES
REGISTERS
-δ t₁+δ
PULSE FROM SIGNALING CIRCUIT
"E"
RAMP GENERATOR
t₀
-δ t₁ +δ
JITTER DUE TO NOISE

Fig. 5. Interval timer block diagram

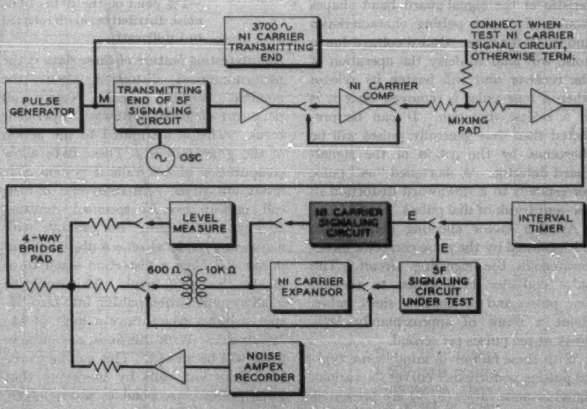

3700 ∿ NI CARRIER TRANSMITTING END
CONNECT WHEN TEST NI CARRIER SIGNAL CIRCUIT, OTHERWISE TERM.
PULSE GENERATOR
M
TRANSMITTING END OF SF SIGNALING CIRCUIT
OSC
NI CARRIER COMP
MIXING PAD
4-WAY BRIDGE PAD
LEVEL MEASURE
NI CARRIER SIGNALING CIRCUIT
E
INTERVAL TIMER
E
600 Ω
10K Ω
NI CARRIER EXPANDOR
SF SIGNALING CIRCUIT UNDER TEST
NOISE AMPEX RECORDER

Fig. 6. Arrangement for signaling circuit tests

cedure used in this study comparing signaling circuits.

Testing Procedure

For the noise measurements, an LDR is used as discussed. To measure signaling distortion, an interval timer is used. This device, developed at Bell Telephone Laboratories by O. R. Miller, generates an accurate linear ramp function. Using this, a translation of pulse duration into an amplitude function is obtained which is then detected by a group of biased amplitude gates. The gates in turn trigger a bank of registers which present a cumulative distribution curve of deviation in pulse duration or per-cent break output from a signaling circuit. Fig. 4 is such a distribution. A block diagram of the timer is shown in Fig. 5. The instrument design permits the use of different time bases, different time interval separations between registers, and a centering adjustment such that the zero distortion register may be set at any pulse duration or per-cent break output desired.

With these devices as the primary measuring instruments, a test setup shown in block diagram on Fig. 6 was assembled. The arrangement consists of a stable source of pulses variable in per-cent break and speed; an Ampex *400A* tape recorder as a noise source, a 4-way bridging pad for mixing and level measurement purposes, the single-frequency signaling circuit under test, the interval timer for pulse distortion distribution data, and appropriate gain and loss to adjust system levels. Provision is also made to insert an *N*-carrier channel unit voice-frequency section in the transmitting and receiving paths to check the 3,700-cycle signaling circuit of this carrier system and to per-

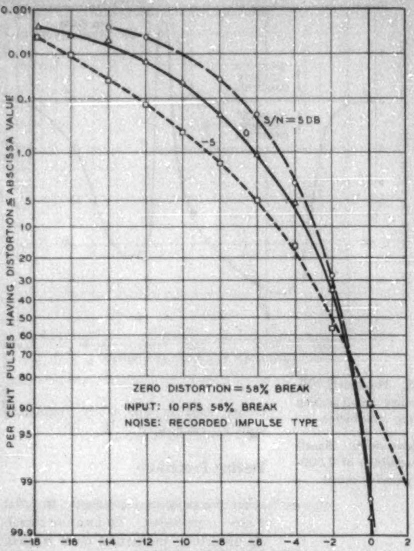


Fig. 8. 2,600-cycle signaling circuit pulsing performance

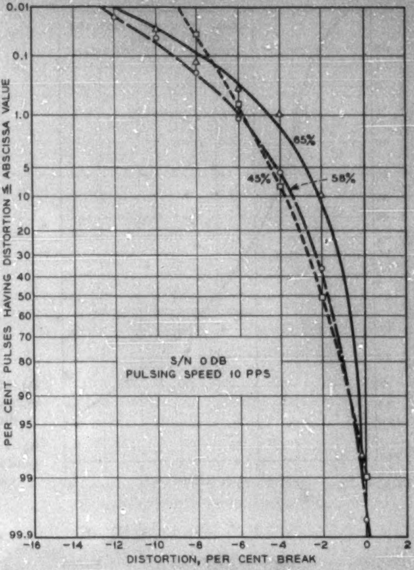


Fig. 9. 2,600-cycle signaling circuit, pulsing performance for different input on–off ratios

Vol.
Pt. 1
1956

mit evaluation of the effect of the *N*-carrier compandor on the signaling-noise problem.

Performance of 2,600-Cycle Signaling Circuit

The 2,600-cycle signaling system is the current standard in-band toll signaling circuit in the Bell System. Some special characteristics of this circuit should be understood, however, before pulsing performance can be discussed. Since this is an in-band signaling circuit, it is necessary to prevent signal imitation by speech. This is accomplished by providing a guard circuit which is responsive to all voice frequencies except the signal frequency, a signal circuit which is responsive to the signal frequency (2,600 cycles), and a detector which is responsive to the difference between the signal and the guard circuits. Speech often contains components of signaling frequency but nearly always contains components in the guard band at the same time. The signal-guard responses are adjusted so that no false operation is obtained for this case. The response of the signal and guard channels of the 2,600-cycle signaling circuit is shown in Fig. 7 taken after Weaver and Newell.[1] When a composite wave of signal plus noise is impressed on this circuit, the noise in the wide guard band will tend to prevent the circuit from responding while the signal and noise in the signal band will act in the direction to operate the receiver. It would be expected then that the particular characteristic of the signal-guard band shapes would affect the pulsing characteristics and this is the case. Guard voltage due to noise will tend to delay the operation of the receiver and will hasten its release when a tone pulse is removed since it is in a release direction. It can be predicted then that generally pulses will be shortened by the action of the signal-guard detector. A shortened "on" pulse corresponds to a downward distortion in per-cent break of dial pulses.

Another factor affecting performance is introduced by the pulse corrector incorporated in the signaling circuit. The pulse corrector is designed to shorten long pulses and to lengthen short pulses about a mean of approximately 58% break at ten pulses per second.

With these factors in mind, some typical pulsing performance curves for various signal-to-noise ratios (S/N) are presented in Fig. 8. The ordinate is per cent of pulses having distortion equal to or less than the abscissa value of distortion in per-cent break below 58% at ten pulses per second. Each curve is labeled with applicable S/N ratio defined as:

$$S/N \text{ in db} = \begin{array}{l}\text{2,600-cycle signal level in db}\\ \text{referred to 1 milliwatt}\\ -1\% \text{ point on the 10-ms LDR}\\ \text{noise distribution in db referred}\\ \text{to 1 milliwatt}\end{array}$$

An interesting feature of these data is the nonsymmetrical distortion about the nominal per-cent break. Nearly all pulses are distorted downward in per-cent break. This is attributed to the action of the guard circuit. These data allow computation of lost calls if system S/N levels are given. Alternatively, if lost-call performance is specified, limiting S/N can be obtained. For example, with a system S/N level of −5 db, 0.04% of pulses will have a distortion equal to or greater than 14% break. This would be a failure point when pulsing into step-by-step switches which have a limit of 44–72% break. With this limit, one pulse in 2,500 will be missed. This can be translated into lost calls by observing that 2,500 pulses correspond to about 2,500/35 = 71.5 calls since an average 7-digit

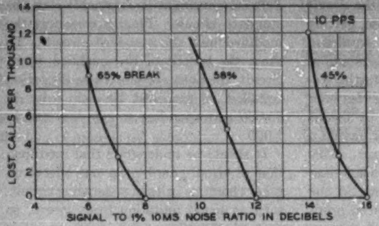
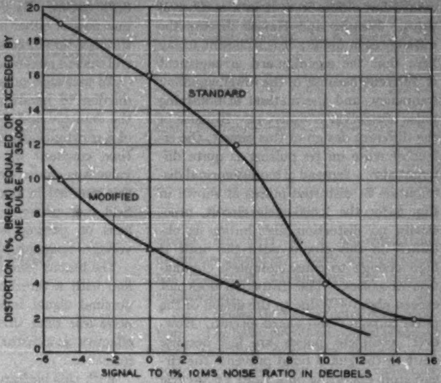


Fig. 10 (above). 1,600-cycle signaling circuit performance with impulse noise

Fig. 12 (right). 2,600-cycle signaling circuit; pulsing improvement with modified signal-guard ratio



number will contain about 35 pulses. This is obviously submarginal performance but a similar calculation for an S/N of $+5$ db gives a lost-call ratio of better than one call in 1,000. This is quite adequate performance, especially since the noise used to obtain the data of Fig. 8 represents that occurring during the most disturbed periods.

Fig. 9 shows the effect of pulse correction on performance. On this curve is plotted a cumulative distribution of distorted pulses against distortion in per-cent break with three different input pulse on–off ratios. Curves are shown for 10-pps (pulses per second) 45% break; 10-pps 58% break; and 10-pps 65% break. In each case in Fig. 9, zero distortion corresponds to the steady output pulse per-cent break without noise. The distortion is measured in percentage points above or below this value. At large distortions, more pulses are distorted for the 58% break condition than for the other two. The lengthening of short pulses and shortening of long pulses is thus apparent.

Performance of 1,600-Cycle Signaling Circuit

The 1,600-cycle signaling circuit was a forerunner of the 2,600-cycle circuit dis-

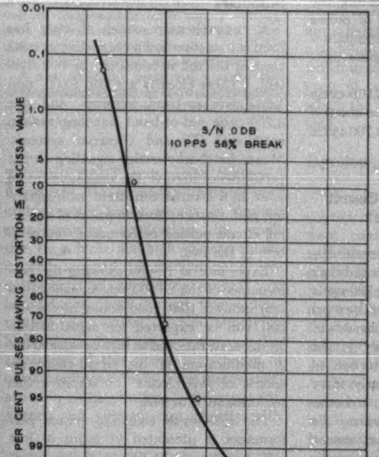


Fig. 11 (left). 3,700-cycle signaling circuit pulsing performance

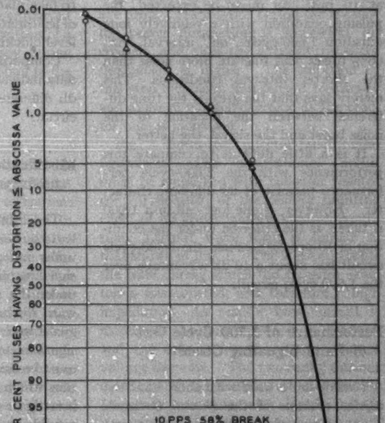


Fig. 13 (right). 2,600-cycle pulsing performance showing effect of compandor cussed but differs from it in a number of ways. Weaver and Newell discuss the differences in detail; it is sufficient to say here that the signal-guard arrangement is different because of the lower signaling frequency, and the method of obtaining pulse correction is different. Because of the different design of this circuit, the effect of noise on its pulsing is quite different also. Instead of a quasinormal distribution of distorted pulses as shown in Fig. 8 for the 2,600-cycle circuit, practically no distortion distribution is obtained on this circuit until the noise is high enough to cause complete elimination of pulses. The distortion does not exceed about 4% due to the action of the pulse corrector and guard circuit. However, complete pulses are lost because noise in the guard band prevents the circuit from operating on a valid 1,600-cycle signal. The test arrangement used with the 2,600-cycle circuit is not adequate to measure the performance of this circuit; nor is the distribution plot of Fig. 8 adequate to specify this performance. Instead, simple pulse counters on the sending leads and receiving leads were installed. The numbers of lost pulses for various S/N levels are then plotted. Fig. 10 is such a plot. Here the number of lost calls per 1,000 is plotted against the signal-to-1%-10-ms-noise ratio. One lost call per 1,000 is defined as one lost pulse in 35,000 pulses. In Fig. 10, data are shown for 45, 58, and 65% break inputs at 10 pps. As might be expected, the pulsing condition with a relatively long duration 1,600-cycle "on" interval, viz., 65% break, can tolerate more noise than the shorter interval conditions. This merely says that the greater the time difference between the duration of the noise burst and the signal, the better.

It is a little difficult to compare this performance with the 2,600-cycle performance because of its different character. However, if one lost call in a 1,000 is taken as the criterion under the condition of 10-pps 58% break, the 2,600-cycle circuit has about a 10-db advantage over the 1,600-cycle circuit.

Performance of 3,700-Cycle Out-of-Band Signaling Circuit

This circuit is used with short-haul toll carrier systems of the *N* and *O* types. This scheme of signaling is simple and straightforward as shown in the block diagram of Fig. 1. Three parts of the circuit need some mention to set the noise performance data in proper perspective.

The band-pass filter has a pass band of about 100 cycles, 3 db, down-centered at 3,700 cycles. By the nature of the circuit any power, either signal or noise, getting through this filter will act like signal and will cause a skewed distortion distribution.

To minimize noise effects, a further reduction in bandwidth is made between the detector and d-c amplifier by means of a resistance-capacitance filter having a time constant of about 12 ms. This value was chosen because delays less than this resulted in poor performance in the field and greater delays introduced problems of pulsing stability with battery voltage, temperature, and time.

The limiter used in this circuit is a flip-flop with a threshold about 6 db below nominal signal level. This means that noise less than this amount will be less effective as a disturbing influence on performance.

With this background, Fig. 11 presents experimental results on pulsing performance of the *N*-carrier 3,700-cycle out-of-band signaling circuit. The most interesting aspect of this plot is that it is skewed in the upward direction of distortion. The effect of noise is generally to increase the per-cent break of pulses. This is expected because noise occurring at the end of a pulse will give the impression that the signal pulse remained just a little longer than it actually did. Similarly, noise occurring in the tone-off interval just prior to the pulse will look like a legitimate signal anticipating the regular pulse. For shortening of pulses to take place there must be a cancellation of legitimate signal by noise, or some high-level blocking must occur.

Comparison of 2,600- and 3,700-cycle data indicates that there may be about 6 db difference in favor of the 3,700-cycle circuit.

Modification of 2,600-Cycle Circuit for Improved Performance

The 2,600-cycle signaling circuit distorts pulses downward in per-cent break under noisy conditions; the 3,700-cycle signaling circuit distorts pulses upward under the same conditions. The downward distortion is due primarily to the guard action; the upward distortion to signal imitation by noise. Clearly there must be some optimum signal-guard relation which maximizes performance. To investigate the amount of improvement possible, a 2,600-cycle circuit was modified experimentally until best performance was obtained. The modification of course effectively reduced the guard action just sufficiently so that the average signal imitation by noise was canceled by guard action. Approximately, a 6- to 8-db improvement is obtainable over the standard circuit. Fig. 12 compares the modified and the standard circuit. To show the benefit of the modified guard circuit, the data are plotted as distortion equalled or exceeded by one pulse in 35,000 corresponding to one lost call in 1,000.

It should be understood that modification of signal-guard ratio has effects on other characteristics of a signaling circuit, such as signal imitation from speech, so that whether this entire improvement is actually obtainable depends on factors other than noise. The present 2,600-cycle signaling circuit has worked well in the toll network for a number of years so it is not proposed to modify existing circuits. Future developments, however, will make use of this signal-guard optimum design.

Compandor Action on Signaling

On *N*- and *O*-carrier systems, the in-band signaling systems must send signaling tone through the system compandor. A subsidiary result of these tests presents data showing that the action of the compandor on signaling is minimal. Fig. 13 shows typical data for the 2,600-cycle system. The differences obtained between the two conditions are insignificant.

Summary

A comprehensive series of tests has been run on standard a-c signaling circuits used in the toll network of the Bell System. These circuits are the 2,600- and 1,600-cycle in-band systems, and the 3,700-cycle out-of-band signaling system used with *N*- and *O*-carrier systems. Means for characterizing signaling circuit performance involve the measurement of noise in a manner consistent with signaling performance; measurement of signaling circuit pulsing performance; and relation of the two.

Experimental results showing the performance of the 2,600-cycle signaling system indicate that good action of the circuit can be expected for signal-to-1%-10-ms-noise ratios of as high as about 0 db. A modification of the signal-guard response of this circuit can improve this performance 6 to 8 db.

The 1,600-cycle signaling system performance is presented in terms of lost calls because of the nature of the circuit action under noise conditions. This circuit can be expected to be about 10 db more sensitive to noise than the 2,600-cycle system.

The 3,700-cycle out-of-band system used on *N*- and *O*-carrier systems has something less than 6-db advantage over the present standard 2,600-cycle system and about equal to the modified design. Pulses are distorted upward in per-cent break generally for this system.

N- and *O*-carrier compandor action on in-band signaling circuit performance in the face of noise is not significant.

References

1. In-Band Single-Frequency Signaling, A. Weaver, N. A. Newell. *Bell System Technical Journal*, New York, N. Y., vol. 33, Nov. 1954.

2. A Signaling Circuit for N-1 Carrier, L. A. Weber. *Bell Laboratories Record*, New York, N. Y., Apr. 1953.

3. Frequency Shift Telegraphy—Radio and Wire Applications, J. B. Davey, A. L. Matte. *AIEE Transactions*, vol. 66, 1947, pp. 479–94.

4. Frequency-Shift Signaling Circuit for 45-Type Carrier Systems. K. E. Appert, R. S. Caruthers. *Ibid.*, vol. 74, pt. I, Nov. 1955, pp. 613–19.

5. The Type N-1 Telephone Carrier System, R. S. Caruthers. *Bell System Technical Journal*, New York, N. Y., Jan. 1951.

6. Sound Level Distribution Recorders, H. Kahl. *Bell Laboratories Record*, New York, N. Y., vol. 17, Apr. 1939.

Magnetic Regulation Transistor Power Supply

LAMBERT F. LYONS
NONMEMBER AIEE

Synopsis: The switching-transistor saturable-reactor d-c to a-c converter is often used to supply alternating current or high-voltage direct current from a low-voltage d-c source. In this application the output voltage varies directly as input voltage, and also varies with load in accordance with the power supply output impedance.

In output regulation of this supply it is usually essential to retain the high efficiency of the basic circuitry. This paper proposes a method of regulation which employs saturable reactors in the input or load circuitry of the power supply. The method provided is convenient for achieving output voltage compensation with supply voltage drift and variations in load current. For closer regulation of output voltage, a silicon diode voltage reference may be utilized.

A method is also presented whereby the output frequency may be controlled magnetically in a-c motor and magnetic-amplifier applications.

IN AIRCRAFT and missile electronic systems, the switching-transistor–saturable-reactor power supply is finding ever-increasing application. Its small size, ruggedness, and high efficiency make it well adapted to medium-power requirements of alternating current, or high-voltage direct current from a low-voltage d-c source. When supplying high-voltage direct current, a major disadvantage of this circuit is variation of output voltage directly with supply voltage. In a-c applications both output frequency and voltage vary directly with supply voltage. Output voltage also fluctuates with load current because of transistor, transformer winding, and rectifier resistance.

It is often desirable to regulate the d-c output voltage or the a-c output voltage and frequency of such a power supply. In systems where this type of power supply is used, space, weight, and transistor power-handling capacity are limiting factors. Therefore, efficiency is of major importance. A variable-series dissipative element employed for regultion is thus not satisfactory. It was determined that a switching-type device must be employed which would vary the percentage time of a cycle of power supply frequency during which the output voltage is applied to the load.

Development

Two types of switching devices for series control of power supply input or output were considered in this investigation: transistors and saturable reactors. In applying the switching transistor to the power supply output, the transistor voltage rating was a limiting factor. For this reason, and the fact that saturable reactors are readily controlled, it was decided to develop a magnetic regulator.

Three configurations were considered: reactors in series with the transistor bases, reactors in series with the transistor collectors, and reactors in series with the output rectifiers for d-c supplies. In each of the three circuits, current flows through the reactor during half of the cycle, driving it to saturation. During the next half-cycle, current is essentially switched off and the reactor can be reset by a control current. In this respect the circuit functions as a self-saturating magnetic amplifier. Each configuration has particular advantages for certain applications.

Fig. 1 shows the base reactor configuration. The basic circuit employed is somewhat modified from the conventional power supply, in order that the collectors might be grounded for more efficient heat dissipation. Resistors *R1* and *R2* are base current-limiting resistors. Resistor *R3* provides a forward bias on one transistor, to insure starting. The power in the base circuit is comparatively low, requiring small reactors for control, and providing high control sensitivity. This sensitivity is lowered somewhat by base current under transistor cutoff condition. The output wave form of the circuit is as shown. The step at the beginning of each output half-cycle represents the time during which the base reactor is unsaturated and causes low base current and consequently a large transistor voltage drop. During this period the transistor power dissipation is somewhat increased, and thus the average transistor dissipation is higher. The duration of this step increases, and output frequency and average voltage decrease as the reactor bias is increased.

Fig. 2 shows the collector reactor configuration. The conventional circuit must be employed with collector reactor control because of the reactor shunting effect of *R3* under transistor cutoff condition. Output wave form is similar to that shown in Fig. 1. The step represents the time during which the collector reactor is unsaturated, and therefore a high impedance. As reactor bias current is increased, the step increases in duration and the output frequency and average voltage decrease. The two configurations described afford a means of maintaining frequency constant with input voltage variations in a-c power applications.

Fig. 3 shows a high-voltage d-c supply with reactors in series with the rectifiers. The frequency and wave form on the secondary of the transformer is a square wave, and is not appreciably altered by reactor bias. The voltage wave form

---

Paper 56-756, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 20, 1956; made available for printing June 5, 1956.

Lambert F. Lyons is with the Bendix Aviation Corporation, North Hollywood, Calif.

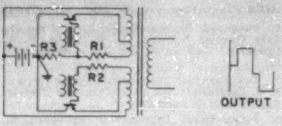


Fig. 1. Base control

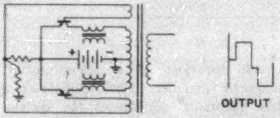


Fig. 2. Collector control

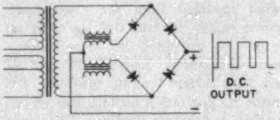
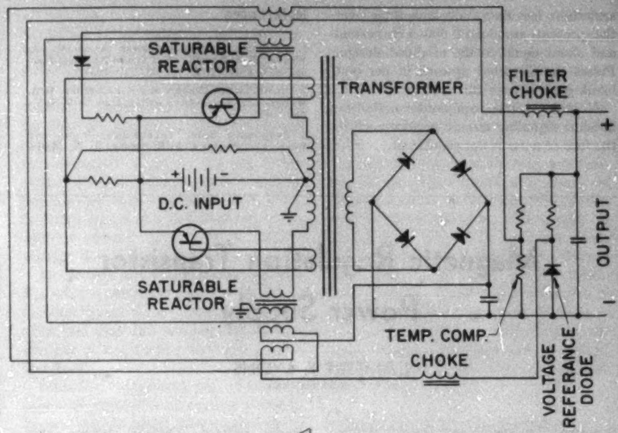


Fig. 3. Output control



9

Fig. 4. D-c output transistor power supply employing magnetic regulation

across a resistive load, however, is a square wave with the leading edge of each half-cycle delayed. The duration of this period of essentially zero voltage varies with reactor bias, and represents the time during which the reactor is unsaturated and presents a high series impedance. A reactive load, such as a choke input filter, will vary the shape of this wave form considerably. However, the voltage integral which the reactor will absorb with a given bias is still the same.

An advantage of this configuration is that the power supply frequency and output voltage are not changed appreciably by operation of the magnetic regulator. Thus an unregulated supply can be provided from the same transformer.

Design

Basic Power Supply

Some sources for design information on transistor-magnetic power supplies are listed in the references. Detailed basic power supply design therefore will not be considered in this paper. Reactors employed in any of the three configurations serve only to decrease the output voltage in accordance with reactor bias. In an unbiased condition the reactors remain saturated, and the circuit functions as a basic power supply, with the exception of the voltage drop across the reactor winding resistance, which is very small with well-designed reactors. Thus the power supply should be designed to supply the maximum current required when the d-c input supply voltage is at minimum operational value. To ensure adequate regulation, the output voltage under these conditions should be about 5 per cent above the desired regulated output voltage.

If a filter is employed for d-c output, it must be of the choke input type, in order to prevent high surge currents in the transistors. Current surges resulting from a capacitor input filter would exceed the saturation current of the transistors, greatly increasing their power dissipation. A small input capacitor may be required for better switching action, while the choke input characteristics are still maintained.

Reactors

The reactors employed must be designed to absorb the difference between the desired output voltage and the output voltage E_{outmax}, under conditions of maximum input voltage and minimum load current. The reactor square-wave voltage rating E_r is determined from the following.

Base Control

$$E_r = E_b(1-a) \qquad (1)$$

where

E_b = base driving voltage
a = ratio of regulated voltage output to maximum output

Equation 1 affords a reactor design for preliminary test only, since base reactor control is greatly dependent on output filter and load characteristics. Final reactor voltage rating is best determined by experiment in particular applications.

Collector Control

$$E_r = E_{\underset{\text{max}}{\text{in}}}\frac{1-a}{a} \qquad (2)$$

where

$E_{\underset{\text{max}}{\text{in}}}$ = maximum power supply d-c input voltage

Output Control

$$E_r = E_{\underset{\text{max}}{\text{out}}}(1-a) \qquad (3)$$

The reactor power or "gate" winding in magnetic amplifier terminology is calculated from the following equation:

$$N = \frac{E_r 10^8}{4FB_mA} \qquad (4)$$

where

N = number of turns
E_r = reactor design voltage
F = power supply frequency with maximum supply voltage
P_m = saturation flux density of core in gauss
A = core cross-section area in CM^2

Standard tape-wound cores of Arnold Engineering Deltamax and *4-79* Mo-Permalloy, or Magnetics, Inc. Orthonol and Hi-mu *80* are used. One-mil material is recommended for frequencies as high as 3,000 cycles.

Small cores afford high sensitivity and high unsaturated reactance, but cause high winding resistance and I^2R loss, since the number of turns required for a

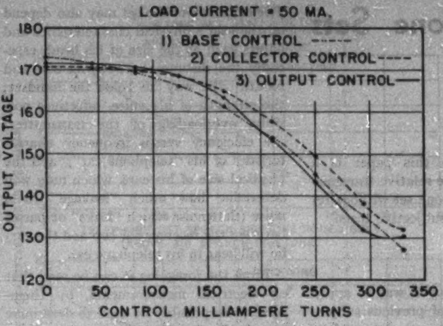


Fig. 5. Reactor control characteristics

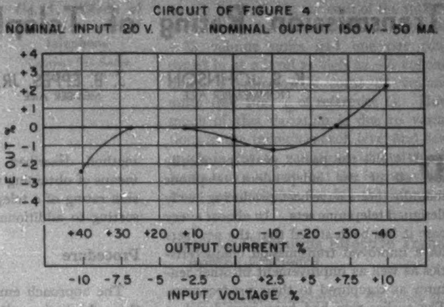


Fig. 6. Voltage regulation without diode reference

given voltage increases with core size reduction. Core selection is in general a compromise between control sensitivity and gate-winding resistance.

Load and Input Voltage Compensation

The gate winding and a test control winding (50 to 100 turns) are placed on the cores. With nominal input voltage and a constant load current, the output voltage versus ampere-turn control curve is plotted. With control windings open, the output voltage versus output current curve is then plotted.

Input voltage compensation is achieved by applying a bias current to the reactor derived from the input voltage. Since this current varies directly with input, there is automatic compensation. If this compensation is not sufficient to make the output voltage essentially constant, the bias current variation may be increased to the desired value by placing a silicon diode of the proper zener voltage in series with the input voltage bias winding and current-limiting resistor. (See Fig. 4.) The number of turns on this winding shall be chosen to maintain the bias current within the zener current rating of the diode. Final adjustment is made by the series resistor.

The required load current compensation can be determined from a consideration of the two curves. These windings are connected in series with the choke for a d-c supply, as shown in Fig. 4. This positive feedback generally is only a few turns, and can be finally adjusted by test. The control winding turns should be adjusted so that, with minimum input voltage and maximum output current, the ampere-turns from the input voltage and the ampere-turns of load current should approximately cancel each other and thus maintain the reactors in a saturated condition. With maximum input voltage and minimum output current, the net effect should be full reactor bias.

Voltage Reference

The control characteristic of the saturable reactors is not linear, and consequently there is some variation of output voltage with input and load. Temperature changes will likewise cause a drift in output.

Final regulation of the order of 1 per cent or better can be achieved by the addition of a silicon-diode reference unit with temperature compensation, as shown in Fig. 4. Additional reactor windings are provided for this purpose. An isolation choke is also provided to minimize alternating current in the control windings.

Performance

The circuit configuration of Fig. 4 was employed for test purposes, and the following components used:

transistors, Sylvania *GT-731*
supply frequency, 1,500 cycles at 20 volts input
silicon reference diodes type *IN203*
rectifiers, two *IN93* germanium and two *IN337* silicon
reactors, Magnetics, Inc. *50176-1D*
filter, 0.015 microfarad, 0.250 henry, 1.5 microfarad
input voltage,–20 volts±10 per cent
output voltage, 150 volts
output current, 50 milliamperes±40 per cent

Fig. 5 shows the control characteristic with an input of 20 volts and a 50-milliampere load. Fig. 6 shows regulation characteristics without the silicon-diode reference unit in the circuit. With the silicon-diode reference unit connected, regulation was 1 per cent. Efficiency at high-current levels was 75 to 80 per cent, a value only slightly lower than that achieved without the regulator. The regulator was also tested, using the configurations of Figs. 1 and 3 (with different reactor gate windings). Control characteristics are included in Fig. 5.

Regulation was adequate in all three test circuits, but increased transistor dissipation was noted in the base control configuration, since operation in the high-dissipation region between transistor saturation and cutoff is necessary during a portion of the cycle.

Conclusions

The transistor–saturable-reactor power supply can be effectively regulated through the use of saturable reactors. The high efficiency, small size, and ruggedness of the basic supply is retained. Regulation is achieved through input voltage compensation, load compensation, and final regulation, using a silicon-diode voltage reference. Either load or input compensation can be employed singly in special applications. In a-c power systems, power supply frequency can be stabilized as well as output voltage.

References

1. A Switching Transistor D-C to A-C Converter Having an Output Frequency Proportional to the D-C Input Voltage, G. H. Royer. *AIEE Transactions*, vol. 74, pt. I, July 1955, pp. 322–26.

2. Transistors as On-Off Switches in Saturable-Core Circuits, R. Louis Bright, C. F. Pitman, Jr., George H. Royer. *Electrical Manufacturing*, New York, N. Y., Dec. 1955.

3. A Design Method for Voltage and Reference Device Using a Silicon Diode, D. A. Burt. National Electronics Conference, Chicago, Ill., 1955.

Transmission Rating of Telephone Sets

K. S. JOHNSON
NONMEMBER AIEE

J. B. EPPES, JR.
MEMBER AIEE

DURING the past 2 years practically all of the independent telephone manufacturers have been producing newly designed telephone sets. In almost every case it has been stated that the new sets have improved transmission characteristics as well as improvement in other features as compared to their predecessors. The earlier sets had been considered as "standard sets" for transmission design purposes by most large operating companies.

Every telephone operating company which was a prospective user of one of these new sets was faced with the problem of evaluating its transmission performance.

Since the transmission design of a telephone system is predicated on the use of telephone sets having a specified transmission rating, it is important that an accurate rating of the set be known. As direct distance dialing spreads to more and more independent operating companies, the importance of industry standards to help achieve proper transmission design of exchange plant will increase. At present, unfortunately, there are no known published industry standards to assist a prospective user in evaluating a new telephone set design. It is the intent of this paper to present a method of determining the transmission rating of a telephone set under various line and room noise conditions, which can be applied to any two telephone sets which are to be compared. A further objective is to present a means of correlating the results with any other method of rating.

The mechanical design features of a telephone set can be readily evaluated but the proper transmission rating is more difficult to determine. Telephone-set transmission ratings are usually determined by a combination of electroacoustical and subjective tests. Both types of tests are costly and time-consuming and it is extremely difficult to obtain results which can be reliably compared. The method described in this paper is a means of obtaining the relative transmission rating of a telephone set without resorting to additional subjective tests.

Procedure

The approach employed was to apply the published results of previous subjective tests to electroacoustical test data obtained from the set manufacturers. By conducting electroacoustical tests simultaneously on both the new design and the earlier "standard set," or some other well-known set, directly comparable data could be obtained. Existing subjective test data could then be applied to the differences in responses between the two sets in order to determine their relative transmission efficiencies.

A specification was developed for electroacoustical testing which was based on American Standards Association standards and the Bell Telephone Laboratories' artificial voice and ear test equipment. The data needed included frequency-response curves of the over-all set showing transmitting, receiving, and sidetone levels when the set was connected successively to typical long and short loops and was energized by means of a conventional battery supply circuit.

Field of Application

The method of rating telephone sets which is described herein is considered to be particularly suited to rural plant. This type of plant is mostly of open-wire construction and involves long high-resistance loops with relatively high line noise and battery supply losses.

General Considerations

The problem of rating a telephone set is a complicated one since the transmission efficiency of the set depends to a great extent on the individual characteristics of the user. Consequently, while electrical measurements alone can help greatly in determining the efficiency of a telephone set, the user's reactions to room and line noises, to acoustic sidetone, and to the over-all attenuation or loss of the circuit are factors which often determine its efficiency in telephone transmission. The efficiency of the set may also depend on the user's physical characteristics and habits such as the size of his head, especially the distance between his ears and mouth; the way he holds the handset; the strength of his voice, affecting, possibly, overloading of the transmitter; the efficiency versus frequency characteristics of his "telephone ear"; and the physical size of his ears, which may well determine how much "leakage" room noise (that noise which "leaks" or passes between the receiver ear cap and the ear) he will hear in his telephone ear.

From the foregoing it can be seen that no electrical measurements, by themselves, can possibly be used to determine completely the efficiency of a telephone set. Such measurements, however, are of great significance if it is necessary to rate only the relative efficiency of a telephone set in terms of some other well-known set which is in widespread use. The Western Electric type *300* set is used in some three quarters of all telephone systems in the United States and so has been taken as the basis of reference in the following quantitative discussion on the efficiency of a specific telephone set.

In developing this procedure an attempt has been made to utilize published subjective test data to the maximum practicable extent.[1–3] Figs. 1 through 3 of this paper are reprinted from references 2 and 3 by permission of the *Bell System Technical Journal.*

When talking face to face, it is generally acknowledged that how loud a person speaks depends on how well he hears himself or how loud his "received" speech (i.e., in both ears) is above that of his "received" noise. That is, his talking level is a function of the total signal-to-noise ratio by which he hears himself. For example, if he is talking in a noisy room, he will tend to speak louder than he will if he is in a quiet room, so that the listener will have the same or a larger signal-to-noise ratio and will therefore hear the talker more easily. When using the telephone, however, the talker does not know, initially at least, whether the listener is understanding him readily or not. If the listener has difficulty in hearing the talker, it may be due to various factors such as (1) a naturally weak voice; (2) a telephone set at the talker's location with a low transmitting efficiency, possibly due to a reduced amount of direct current in its transmitter; (3) a high circuit attenuation or frequency distortion; (4) a large amount of circuit noise; (5) a large amount of room noise at the listener's location; or (6) a

Vol.
Pt. 1
1956

Paper 56-781, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 26, 1956; made available for printing April 27, 1956.

K. S. JOHNSON (deceased) was with and J. B. EPPES, JR., is with the Rural Electrification Administration, Washington, D. C.

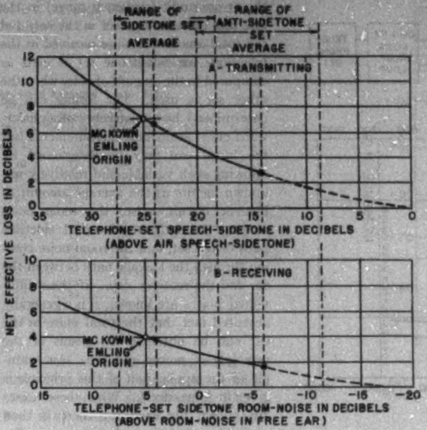


Fig. 1. Effects of sidetone on user of telephone

Courtesy Bell System Technical Journal

telephone set at the listener's location with a low receiving efficiency.

A large difference in the loss of a circuit in the two directions of speech may also make it difficult to predict the eventual level of the talker's speech. Large differences in loss frequently occur when one user is on a high-resistance loop and the other is on a low-resistance loop; they also occur on revertive calls. For example, if *A* is in a noisy room, he is likely to talk initially in a fairly loud voice. If listener *B* is in a quiet room he may hear *A* so well that he thinks the connection is a good one and he therefore tends to respond at a level that is somewhat lower than normal. Speaker *A* in turn then hears *B* poorly and tends to raise the level of his speech still further. This process may continue to build up until comments of poor reception by *A* tend to straighten matters out.

In view of the preceding, it is obviously impossible to determine the level of speech of the average talker with any great degree of precision. Although the average speech level at a local central office is reported[1] to be −19 volume units, the speech level of one per cent of the talkers comes into the central office as low as −33 volume units and one per cent comes in at a level as high as −5 volume units, a variation of some 28 volume units (see Fig. 1 of reference 1). How much of this variation is due to variations in the talker's speech level and how much is due to variations in the circuit losses is not known, but the former variations are probably by far the larger.

Basis for Computing Effective Transmitting Efficiency

The effective transmitting efficiency of a telephone set is improved by:

1. Increasing the acoustical power delivered to the transmitter unit.
2. More efficient conversion, over the audio-frequency range, from acoustical power to electric power at the output terminals of the set.
3. Reducing the acoustical sidetone level in the telephone ear of the talker.

The only known statistical data showing how a talker's speech level varies with sidetone are shown in Fig. 1 (Fig. 5 of reference 2). These data, which were derived from measurements of telephone speech over relatively short loops and tie lines, indicate that for small increases in the acoustical sidetone of an antisidetone set (type *300*) there is an effective loss due to talker's reaction of some 30 to 40 per cent of the amount of the increase. On the other hand, Inglis indicates that if the acoustical sidetone of a telephone

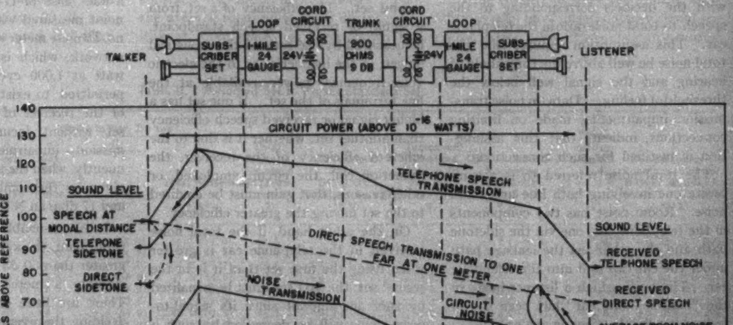


Fig. 2. Power level diagrams of typical telephone connections

Courtesy Bell System Technical Journal

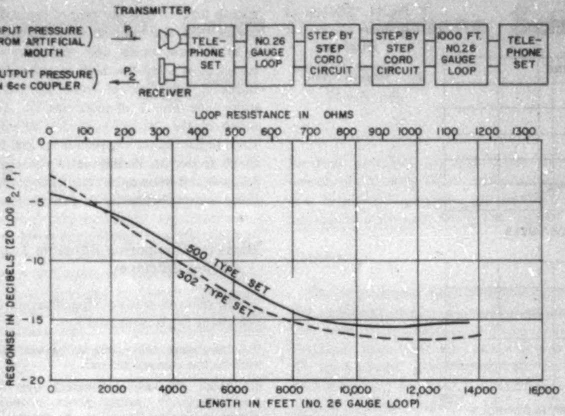


Fig. 3. Comparative sidetone levels

Courtesy Bell System Technical Journal

set under these loop conditions is reduced some 14 decibels (db) or more below that of the type *300* set, the average talker will raise his speech level only about 2.8 db. A sidetone level 14 db below that of the *300* set on a short loop is the level of natural air speech sidetone.

Basis for Computing Effective Receiving Efficiency

The proposed method for comparing the relative receiving efficiency of two sets assumes that the effective efficiency of a set, expressed in decibels, varies directly with the decibels corresponding to the speech-to-total-noise ratio in the telephone ear. This assumption requires that the total noise be well above the threshhold of hearing and the signal well below the threshold of feeling. Data on noise transmission impairments, made on limiting connections, indicate that this assumption is justified for such connections.

The total noise referred to is a composite one involving both line and room noise. Room noise has two components in the telephone ear, one via the sidetone path and the other via the leakage path under the ear cap and into the telephone ear. The noise which a listener hears in the "nontelephone ear" has been shown to have negligible effect on one's ability to receive intelligence in the telephone ear.

The receiving efficiency of a telephone set depends on: its ability to absorb electric audio-frequency power from the line; its ability to convert the absorbed electric power into corresponding acoustical power in the ear, over the audio-frequency range; and its effect on the total noise in the telephone ear.

On an average connection, where the speech-to-noise ratio is large, the impairment due to noise is not directly proportional to the noise increase. As can be seen from Fig. 1, a large increase in sidetone room noise results in only a relatively small net effective loss. However, in situations where the signal-to-noise ratio approaches that of a limiting connection, a one-to-one relationship is assumed to exist between the signal-to-noise ratio of one set as compared to that of the second set. The efficiency of a set, from a signal or received speech standpoint, is measured by the ratio of the acoustical output in the listener's ear to the electric voice-frequency power available at the line terminals of the set. If one set has a higher signal or received speech efficiency than another set, whether it is due to the inherent efficiency of the receiver, the induction coil, the circuit employed, or other reasons, that gain must be credited to the set having the greater efficiency.

On the other hand, if the total noise received in the telephone ear is greater when using the first set than it is in the second set, the first set must be penalized by such an amount since its signal-to-noise value will be decreased by such a fact. Consequently, the problem of finding the relative merits of two telephone sets, from a receiving standpoint, depends essentially upon, finding the relative efficiency of each set in converting the available voice-frequency electric power (over the entire audio-frequency range) in the line into acoustical power in the ear, and the total amount of noise received in the telephone ear when one set is used as compared to the amount of noise when the other set is used. The former can be determined by relatively straightforward electroacoustical measurements but the latter can only be determined by knowing such variable and relatively unknown factors as the average amount of line or circuit noise, the room noise coming through the electrical sidetone path of the set, and the room noise coming through the leakage path between the ear and the receiver cap. If these individual noises are known, it is a generally accepted fact that the total noise in the ear can be obtained by means of the principle of power addition. An example of the application of this principle is given in Appendix I. When these noises are known for each set there is then available all information necessary for rating the relative receiving efficiencies.

The problem thus resolves to determining the average values of the line and room noises that should be assumed to exist in the plant in which the set is to be used. A number of large metropolitan systems assume that a room noise of 40 db rap ("reference acoustic power" which is 10^{-16} watts per square centimeter in a plane wave, and corresponds to a sound pressure of 0.0002 dyne per square centimeter) is a reasonable average value for residential locations and that 65 db rap is a fair average value to assume for business offices. The weighted average value normally assumed is 50 db rap. When there is a room noise of 50 db rap, a line noise of 17 dba (db adjusted, the noise measured with a Western Electric no. *2B* noise meter with the *F1A* weighting network, which is 85 db below 1 milliwatt at 1,000 cycles) is the maximum permitted to exist at the line terminals of the receiver of a standard type *300* set without incurring a noise transmission impairment (NTI). Consequently, when the total noise exceeds that given by the combination of 50 db rap and 17 dba, an NTI is incurred.

It is reasonable to assume that, in rural plant, the average room noise is not greater than 50 db rap and that 40 db is possibly a more representative figure. There are little or no data to help in deciding the average value of line noise in rural plant, but it can be safely assumed that it is materially greater than the average existing in urban and suburban plants because of the relatively large amount of open wire used in rural plant and of the lengths of exposure to power lines.

Table I. Typical Standard Set, Total Noise in Listener's Ear When Using a Standard Set

		Noises in Ear, Db rap			
		A	B	C	
Assumed Room Noise, Db rap	Assumed Line Noise at Receiver Input Dba	Leakage Room Noise	Sidetone Room Noise	Line Noise	Total Noise*
40	17	39	23	41.5	43.5
40	23	39	23	47.5	48.1
40	29	39	23	53.5	53.6
50	17	45	33	41.5	46.8
50	23	45	33	47.5	49.5
50	29	45	33	53.5	54.1
60	17	46	43	41.5	48.7
60	23	46	43	47.5	50.6
60	29	46	43	53.5	54.5
70	17	47	53	41.5	54.2
70	23	47	53	47.5	54.9
70	29	47	53	53.5	56.8

*Obtained by "power addition."

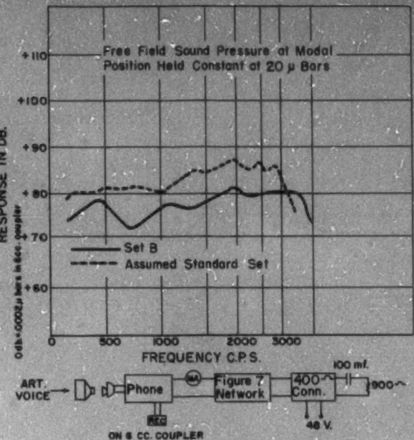


Fig. 4. Sidetone response

The best available data regarding the values of room and line noise in a typical urban connection are given in Fig. 2 (Fig. 8(A) of reference 2). The total room noise for the typical connection which Inglis assumes is made up of three components:

1. The leakage between the ear and the receiver cap.
2. The return noise or noise which is picked up by the transmitter and gets into the listener's receiver by way of the electrical (antisidetone) path of the set.
3. The incoming room noise which is the noise picked up by the transmitter at the talking end of the circuit and is transmitted over the circuit to the listener's receiver.

Unless the circuit has a very low loss or unless the noise in the room at the transmitting end of the circuit is excessively large, the incoming room noise is usually negligible as compared with other room noises. In Fig. 2 the incoming room noise, *D*, is given as about 32 db rap. When power addition is used and when some of the other noises are 40 db rap, or more, this noise is negligible in affecting the total noise in the listener's ear.

The leakage room noise, *A*, depends, among other things, on how tightly the user holds the receiver to his ear. Tests have indicated that for room noises of 45 db rap or more the leakage room noise in the ear varies slowly with changes in the surrounding noise because the listener tends to hold the receiver tighter to his ear when room noise is high. Therefore, the leakage room noise in the ear will be assumed to be substantially constant for these higher noises. For the average room noise of 50 db rap, the leakage room noise has been shown by Fig. 2 to be 45 db rap. In these tests Inglis used an artificial ear and leak. For room noises of 40 db rap or less, it is probable that the receiver is loosely held to the ear so that under such conditions the leakage room noise may be assumed to be only slightly less than that of the ambient room noise. The leakage room noise figures shown in Tables I and II were derived from these considerations.

The power levels shown in Fig. 2 were for a relatively short loop and a large signal-to-noise ratio. In rating telephone sets it is of interest to consider limiting conditions since the limiting conditions may have unsatisfactory signal-to-noise ratios. Fig. 2 assumes an acoustic

Table II. Assumed Set B Versus Standard Set

		Noise in Ear, Db rap									
		A	B	C							
Room Noise, Db rap (Assumed) (1)	Line Noise, Dba at the Receiver (Assumed) *(2)	Leakage Room Noise (3)	Sidetone Room Noise (4)	Line Noise *(5)	Total Noise = n *(6)	Total Noise in Standard Set (From Table I) (7)	Increased Noise in Set B Over Standard Set = N *(8)	Increased Received Speech in Set B = S = r (9)	Gain in Effective Receiving Efficiency = R = S − N *(10)	Transmitting Gain = T t+t₁+t₂ T = 3.0+ 2.5+1.0 (11)	Net efficiency Equals gain of Set B Over Standard Set = (T+R)/2Db *(12)
40	17 (27.4)	39	22.0	43.5 (53.9)	44.8 (54.0)	43.5	+1.3 (+10.5)	+2.0	+0.7 (−8.5)	+6.5	+3.55 (−1.00)
†40	23 (28.4)	39	22.0	49.5 (54.9)	49.9 (55.1)	48.1	+1.8 (+7.0)	+2.0	+0.2 (−5.0)	+6.5	+3.35 (+0.75)
40	29 (31.1)	39	22.0	55.5 (57.6)	55.6 (57.7)	53.6	+2.0 (+4.1)	+2.0	+0.0 (−2.1)	+6.5	+3.25 (+2.20)
‡50	17 (27.4)	45	32.0	43.5 (53.9)	47.5 (54.5)	46.8	+0.7 (+7.7)	+2.0	+1.3 (−5.7)	+6.5	+3.90 (+0.40)
†50	23 (28.4)	45	32.0	49.5 (54.9)	50.9 (55.3)	49.5	+1.4 (+5.8)	+2.0	+0.6 (−3.8)	+6.5	+3.55 (+1.35)
50	29 (31.1)	45	32.0	55.5 (57.6)	55.9 (57.9)	54.1	+1.8 (+3.8)	+2.0	+0.2 (−1.8)	+6.5	+3.35 (+2.35)
60	17 (27.4)	46	42.0	43.5 (53.9)	48.9 (54.8)	48.7	+0.2 (+6.1)	+2.0	+1.8 (−4.1)	+6.5	+4.15 (+1.20)
60	23 (28.4)	46	42.0	49.5 (54.9)	51.6 (55.6)	50.6	+1.0 (+5.0)	+2.0	+1.0 (−3.0)	+6.5	+3.75 (+1.75)
60	29 (31.1)	46	42.0	55.5 (57.6)	56.1 (58.0)	54.5	+1.6 (+3.5)	+2.0	+0.4 (−1.5)	+6.5	+3.45 (+2.50)
70	17 (27.4)	47	52.0	43.5 (53.9)	53.6 (56.6)	54.2	−0.6 (+2.4)	+2.0	+2.6 (−0.4)	+6.5	+4.55 (+3.05)
70	23 (28.4)	47	52.0	49.5 (54.9)	54.7 (57.1)	54.9	−0.2 (+2.2)	+2.0	+2.2 (−0.2)	+6.5	+4.35 (+3.15)
70	29 (31.1)	47	52.0	55.5 (57.6)	57.5 (58.9)	56.8	+0.7 (+2.1)	+2.0	+1.3 (−0.1)	+6.5	+3.90 (+3.20)

*Figures in parentheses are for a 1-ringer unbalance which causes 27 dba of line noise.

† Typical rural plant conditions.

‡ Typical urban plant conditions.

Average effective gain; $(T+R)/2 = +3.76\ (+1.74)$ db.

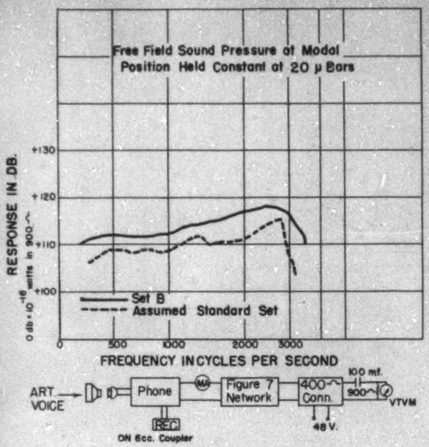


Fig. 5. Transmitting response

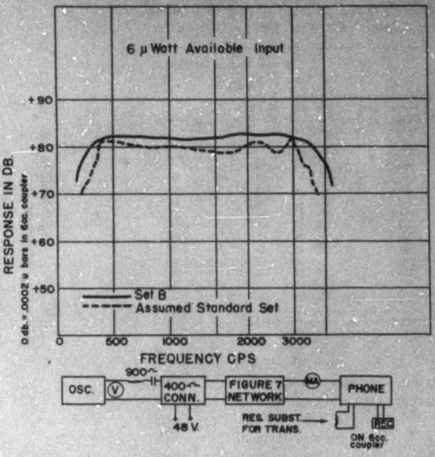


Fig. 6. Receiving response

sidetone loss of 10 db (i.e., the difference between average room noise of 50 db and a return loss, *B*, or 40 db.

For long-loop conditions which are of the greatest interest in rating a telephone set, the acoustic sidetone loss may be appreciably greater than the 10 db shown by Inglis, partly because of the greater battery supply loss of the transmitter on long loops and the anti-sidetone characteristic of the set. Data on the sidetone response of a telephone set with a typical long loop are part of the electroacoustical data obtained from set manufacturers. The way in which sidetone level varies with the loop for the Western Electric type *300* and *500* telephone sets is shown in Fig. 3 (Fig. 6 of reference 3). It should be noted that for loops of about 800 to 1,200 ohms resistance, the sidetone loss of the type *300* set is about 17 db. The fictitious standard set discussed herein has an average sidetone loss of about 17 db on a typical long loop, as shown in Fig. 4. The difference in sidetone loss for the two sets under consideration can be determined by comparing the sidetone response curves.

Several values of room and line noises were assumed in columns 1 and 2 of Table I for a typical standard set. The values in column *A* were obtained as indicated. Column *B* was obtained by considering the sidetone room noise to be 17 db below the room noise. Column *C* is column 2 converted from decibels adjusted at the receiver input to decibels rap in the telephone ear. To do this, 45 db were added to the decibels adjusted value to convert from the decibels adjusted reference to the decibels rap reference of 10^{-16} watts and 20.5 db were subtracted to allow for the loss from the receiver input to the ear (see Fig. 2, 58.5 db input to receiver and 38 db input to ear.) The last column of Table I shows the total noise in the ear, and was obtained for each combination of line and room noise by summing the corresponding values of columns *A*, *B*, and *C*, by the power addition process.

Figs. 4–6 show typical frequency-response curves for a fictitious set *B* and a fictitious standard set. These curves were assumed to have been obtained when the sets were connected through the artificial line network shown in Fig. 7 and a typical battery supply circuit. The transmitting response *t* of set *B* is shown by Fig. 5 to be about 3 db better than that of the standard set. Fig. 4 shows the sidetone response of set *B* to be about 6 db lower than that of the standard set. Fig. 6 shows set *B* to have an average gain of about 2 db in receiving response *r* over that of the standard set. The Fig. 7 network is equivalent to a typical rural line consisting of 3.41 miles of no. 22 gauge *H44* loaded cable circuit, and 5.11 miles of no. 109 steel-line-wire circuit. The loss in this network is shown in Fig. 8. The handset handle of set *B* is also assumed to be 1/2 inch shorter than the handle of the standard set. It is assumed that shortening of the handset handle results in a 2.5-db transmitting gain t_1 due to the closer coupling between the talker's lips and the transmitter unit.

With these assumptions and with the aid of Fig. 1, the effect of acoustic sidetone on the transmitting efficiency was determined. As previously stated, the sidetone level of the standard set on a typical long loop was assumed to be 17.0 db below the transmitting level, or 7.0 db below the type *300* set average indicated in Fig. 1(A). This level is 7.0 db above air speech sidetone and results in a transmitting talker's reaction penalty of 1.0 db. Set *B* would have a sidetone level 1.0 db above air speech sidetone and a transmitting talker's reaction penalty of 0.0 db. The net transmitting gain of set *B* over the standard set due to talker's reaction $= t_2 = 1.0 - 0.0 = 1.0$ db. Fig. 1(B) was not used in this rating method since the effect of sidetone as well as other noise components on receiving efficiency was included in the summation of the total noise in ear.

Tests have been made to show the effect of a 1-ringer unbalance on a multiparty line which had a moderately severe exposure to power-line induction. The results of these tests indicated that if a line is equipped with high-impedance multifrequency ringers, a 1-ringer unbalance may cause about 27 dba of noise. Table II shows how fictitious set *B* compares with a standard set. The effect of the addition of 27 dba of noise is shown by the numerals in parentheses.

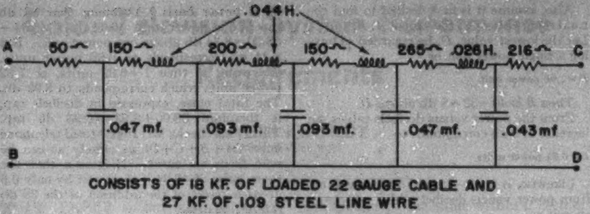


Fig. 7. Network for typical long loop

Except for these numerals it can be seen that columns 1, 2, and *A* of Table II for set *B* are identical to the corresponding columns in Table I. Column *B* of Table II was obtained by starting with the corresponding value in column *B* of Table I, adding 3 db for the more sensitive transmitter of set *B*, subtracting the 6-db increased sidetone loss of set *B*, and adding the 2-db increased receiver efficiency of set *B*. The first figure in column *B* of Table II therefore is $(40-17)+3-6+2=22.0$ db. Column *C* and the total noise for set *B* in Table II were obtained in the same manner as in Table I. The increased noise in set *B*, column *N*, was obtained by taking the difference between the total noise in the ear for set *B* and the total noise in the ear for the standard set. The received signal *S* is 2 db higher for set *B* due to the inherently more efficient receiver in set *B*. The effective gain in receiving efficiency $R=S-N$ and is shown in column 10. Column 11 shows the transmitting gain *T*, which is made up of the gain due to the more efficient transducer, *t*; the gain due to the shortened handset handle t_1; and the gain due to less talker's reaction t_2. The last column shows the net effective gain of set *B* over the standard set, expressed in the customary manner as $(T+R)/2$.

Column 10, Table II, illustrates how easily and completely a gain in the efficiency of a receiver as a transducer could be nullified by an increase in line noise such as would be caused by a 1-ringer unbalance. On the other hand, all the gain in the efficiency of a transmitter as a transducer would be effective, provided it is not accompanied by an increase in the acoustic sidetone level, or less efficient coupling between the talker's lips and the transmitter. The effect of increased line noise, of course, would be most pronounced where the line noise was the controlling noise.

From the foregoing it is apparent that the relative transmission efficiencies of any two telephone sets can be determined for any assumed values of room and line noises provided that the electroacoustical transmitting, receiving, and sidetone frequency-response curves of the two sets are known.

First, the total noise in the telephone ear must be determined by the power addition of the leakage room noise, the sidetone room noise, and the line noise. To perform this addition it is convenient to express each of these three noise components in decibels rap at the listener's ear for each of the two sets being compared. The difference in the two totals gives *N*, the net increased noise of one set over the other.

Next, the receiving frequency response curves are compared to determine the difference of one set over the other in receiving efficiency $(r=S)$. The gain in effective receiving efficiency is then taken to be $R=S-N$. The transmitting frequency-response curves are then compared and the differences between them, *t*, is corrected by the net effect of coupling between the talker's lips and the transmitter t_1, and the effect of talker's reaction t_2 due to sidetone, as obtained from Fig. 1. The transmitting gain $T=t+t_1+t_2$.

The net effective gain of one set over the other is then $(T+R)/2$ for the particular line and room noises assumed. A different value would be obtained for each different value of assumed line and room noise.

Conclusions

It is believed that the method of transmission rating of telephone sets described in this paper provides a valuable tool for use by operating telephone companies in evaluating the performance of telephone sets in their systems. It does not take the place of an industry standard method of rating, but it could serve as the basis for the development of such a standard. It is not based on arbitrary standards but on the results of subjective tests. If later subjective tests result in modifications, the method of rating can still be applied without change.

Appendix I. Power Addition of Noises

It is generally accepted that the resultant or total noise, which is composed of two or more noises of different composition, is that which corresponds to the total power of the individual noises. For example, suppose

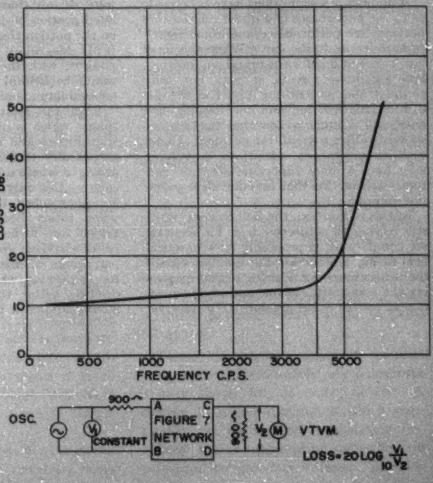


Fig. 8. Loss of network

that there are two such noises of equal power. The total power is obviously twice as great as is that of either of the individual noises. Expressed in decibels, a doubling of the power corresponds to an increase of 10 $\log_{10} 2$ or of 3 db. Hence the resultant noise in the case of two such equal powers is 3 db greater than that of either individually.

This process means that if several sources of noise exist at a given place and if one of the component noises is several decibels louder than that of any of the others, the total noise, expressed in decibels, is only slightly greater than that of the loudest of the component noises.

As a numerical example, suppose that in the receiver of an electrically short connection (Fig. 2) the component room noises are

A—Leakage under the receiver cap....	45 db rap
B—Return room noise, through the sidetone path....................	40 db rap
D—Incoming room noise, from the distant end of the circuit...........	32 db rap
C—The circuit noise at the ear*.......	38 db rap

Also, assume it is first desired to find the total room noise, i.e., the sum of A, B, and D. Let the lowest value, D, be regarded as having a single power unit or

$D = 1.00$ power unit

Then B is $40-32=8$ db above D.

From the power versus decibels tables, an increase of 8 db corresponds to

$B = 6.31$ power units

Likewise, A is $45-32=13$ db above D or, from power versus decibel tables, A corresponds to

$A = 19.95$ power units

The total number of power units in $A+B+D$ is then 27.26.

A power ratio of 27.26 corresponds to a gain of 14.34 db. Hence, the total room noise $A+B+D$, is 14.34 db above 32 db rap or is $14.34+32.00$ or 46.34 db rap.

The circuit or line noise assumed, C, was 38 db rap. In order to determine the combined or total room noise and circuit noise, the two noises must also be added together on a power basis. Assuming that 38 db rap is one power unit, 46.34 db is $46.34-38.00=8.34$ db above one power unit. But 8.34 db corresponds to 6.82 power units. The total is then $1+6.82$ units, or 7.82 power units which corresponds to 8.93 db. The total noise, expressed in decibels rap, is therefore $38.00+8.93=46.93$ db rap. This value checks Mr. Inglis' total telephone noise $(A+B+C+D)$ as closely as can be read from his curve. It is seen that the room noise of 46.34 is increased by only 0.6 db as a result of the addition of the 38 db rap of line noise.

---

* Actually this circuit noise can be increased by about 3.5 db, to 41.5 db rap without causing any NTI.

References

1. Speech Volume Survey on Telephone Message Circuits, V. Subrizi. *Bell Laboratories Record*, New York, N. Y., Aug. 1953.

2. Transmission Features of the New Telephone Sets, A. H. Inglis. *Bell System Technical Journal*, New York, N. Y., July 1938.

3. An Improved Telephone Set, A. H. Inglis W. L. Tuffnell. *Ibid.*, Apr. 1951.

◆

Discussion

W. H. MacAdam (American Telephone and Telegraph Company, New York, N. Y.): The paper discusses an interesting approach to the important but difficult problem of comparing the performance of different types of telephone sets. While the method is useful in indicating qualitative trends with changes of one or another of the many variables, the difficulty seems to be in obtaining quantitative agreement with transmission performance as indicated by observations and tests on actual users. Perhaps further study will develop means to produce comparable results.

A number of assumptions have been made as to the importance of a direct relationship between set performance and total signal-to-noise ratio at the ear. While this may not be out of line for the marginal, limiting-loss high-noise circuit, it may not fully consider the present-day user's desire for comfortable volume and for keeping line noise and sidetone so low that they do not seriously affect normal transmission. Under such circumstances, variations in volume may have a more important effect on customer satisfaction than an equivalent variation in noise.

In this connection, the Bell System, while still relying on subjective tests for accurate set comparison, is proposing as a practical field-rating system for the over-all circuit the measurement of acoustic pressure entering the transmitter at one end of a connection as related to the pressure at the receiver at the other end. The decision to proceed along these lines is based on the fact that the modern telephone set has a rather flat transmission frequency characteristic, and its sidetone loss is sufficiently high to keep noise or speech sidetone from becoming important.

K. S. Johnson and **J. B. Eppes, Jr.**: Mr. MacAdam indicates that transmission ratings obtained by using this method of rating may not agree with the results obtained by observations and tests on actual users. We based our method on A. H. Inglis' subjective test results (see reference 2 of the paper). The sidetone levels we considered were so low that they fell on the extrapolated portion of Inglis' curves rather than on the portion that represented actual data. If Mr. MacAdam has unpublished data that disagree with the published data used, it would be helpful to make them available for evaluation and comparison.

With the relatively flat frequency response which is obtained from present-day transmitters and receivers, we believe that satisfactory transmission will normally be achieved where a large signal-to-noise ratio exists. It is only when limiting connections are approached that noise and signal-to-noise ratios become important. For a typical low- or medium-resistance loop in a rural telephone system involving a sizeable percentage of noncable facilities, it is not too important whether one set is several decibels better than another set. A set having automatic transmission equalization, such as the Western Electric type *500* set, reduces the transmission improvement achieved by more efficient transmitter and receiver units on a low- or medium-resistance loop.

We concur with Mr. MacAdam regarding the desirability of keeping a comfortable volume and a large signal-to-noise ratio. Unfortunately, these conditions are not achieved on a large number of rural telephone circuits with long open-wire exposures to power systems.

We also concur with the desirability of subjective tests for accurate comparison of different telephone sets but feel that it is impracticable to require either the independent manufacturers or small operating companies to conduct such tests. We do not believe that measurement of the difference between the acoustic pressure at the transmitter end of a circuit and the pressure at the receiver end of the circuit would, by itself, give a valid rating of a set, under the variable noise conditions prevalent in rural telephone systems.

We would like to re-emphasize that the rating procedure described in this paper is a practical tool to enable a prospective purchaser of a telephone set to evaluate its transmission performance without resorting to subjective tests. The checks we have been able to make on its accuracy are limited, but have given us reasonably good correlation with the results of subjective tests that were available to us. We would welcome the opportunity of checking this method against the latest available subjective test data.

Vol.
Pt. 1
195

Color-Television-System Performance Requirements

RALPH KENNEDY
NONMEMBER AIEE

THIS PAPER is intended to cover the basic requirements which a transmission system must meet to handle a standard color-television signal. Additionally, where possible, the effects produced by departures from the ideal by various parameters will be illustrated. Three broad areas which will be covered include phase, amplitude, and transient requirements.

Phase Requirements[1]

The placing of phase at the top of the list of requirements is no accident. It actually is the most important parameter which must be considered. Some indication of the effect which phase has may be quickly realized when it is pointed out that three different types of phase relationships must be held to very close tolerances to assure good color pictures.

1. The phase relationship between the high- and low-frequency signal components must be held within ±0.05 μsec (microsecond) which is ±62 deg (degrees) at 3.579 mc (megacycles) per second.
2. The phase relationship between any color component and burst must be maintained within approximately 10 deg.
3. The phase relationship between the I and Q vectors must be 90 deg ± 10 deg. Consider each in detail.

TIME DELAY

Item I has several terms used to describe it. Perhaps the easiest way to visualize the requirement is to use the transmission time approach. It is required that the time necessary for high-frequency signal components to traverse the system be the same within a certain tolerance as is required for the low-frequency components. More precisely, all signal components having frequencies between 50 kc and approximately 5.0 mc must have identical transmission times within certain tolerances. See Fig. 1.

Needless to say, abrupt changes in the phase characteristic are to be avoided since these represent regions of the frequency spectrum where the time of transmission of certain components is markedly different from that of near-by frequencies. Indeed, the maximum extent of such discontinuities has been written into the specification and can not exceed ±0.1 μsec in certain portions of the video band. This region extends from 50 kc to 3 mc. The permissible variations decrease linearly to ±0.05 μsec at 3.579 mc above which they increase linearly to ±0.1 μsec at 4.18 mc.

Fig. 2 shows the desired phase curve.

ENVELOPE DELAY

The standard adopted for the color-system transmission-phase specification is stated in terms of the envelope or group delay characteristic.[2,3] Mathematically, envelope delay is expressed as the derivative of the φ versus w characteristic. Thus, if $\varphi=wt$, $d\varphi/dw=w\,dt/dw+t$. For an ideal system or one where t remains constant for all values of w in the domain of interest $d\varphi/dw=t$. See Fig. 3. However, when t is not a constant $d\varphi/dw$ is the sum of two terms, one of which embodies a rate of change of transmission time with respect to angular frequency. It is this term which justifies the use of envelope delay as a criterion for system performance. The derivative has the effect of magnifying the variations which are in Fig. 1 or 2. This means that the system is being subjected to a much more stringent test when it is required that the envelope delay have a particular tolerance. Conversely, a system which meets a given envelope delay requirement will have a much better appearing phase or simple time delay performance.

As is true for many physical phenomena there are drawbacks to envelope delay measurements.[4] Any realizable network has a relative φ versus w characteristic (see Fig. 2) which has a value of $\varphi=0$ or $n\pi$ radians at $w=0$ where n is a positive integer. Clearly, when $w=0$ one cannot take the derivative of $\varphi=wt$ and there results an undefined point where it is essential that the characteristic have a value.

An example will serve to illustrate. The relative phase characteristic of the side-band filter in a television transmitter is shown in Fig. 4. It rises rapidly from $\varphi=0$, $f=0$ to a frequency of about 1 mc above which it has a smaller slope in the remainder of the band. Envelope delay measurement would appear to be much better above 1 mc than below it and phase correction might appear desirable in the region below that frequency. However, prolongation of the curve back to $f=0$ (curve A) to result in φ having an integer value of π radians is remote indeed. Instead, the lower frequency portion should be considered distortionless and correction applied to the upper portion B.

What kind of a test signal will give a measure of the envelope delay characteristic? As has already been indicated, envelope delay is the angular frequency rate of change of the transmission time at the angular frequency. Since 2π is a constant, envelope delay may be written $d\varphi/df$. Now, since it is known that $\lim_{\Delta f\rightarrow 0} \Delta\varphi/\Delta f=d\varphi/df$ it becomes evident that the envelope delay is a measure of the incremental change in phase occasioned by a change of frequency as the frequency increment approaches zero in the limit. What is ideally necessary is a phase detector capable of measuring incremental changes of phase as the frequency increment becomes smaller and smaller, finally approaching zero as a limit. Further, such measurements must be made at a sufficient number of frequencies to be able to define a curve such as Fig. 3.

At this point it is necessary to depart from mathematical rigor and consider circuit practicalities. Since the phase detector cannot be expected to meet the foregoing dictates, it is necessary to ease the requirements by simply stating that it must indicate some angle $\Delta\varphi>d\varphi$ for $\Delta f>df$.

These assumptions make it possible to use conventional techniques to generate a test signal for envelope delay.[5] A doubly balanced modulator which suppresses the carrier and modulating frequency thereby generating only the two side-band components suffices. The dif-

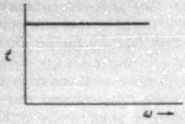
---

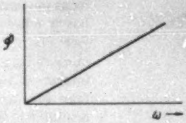
Paper 56-271, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 30–February 3, 1956.

Manuscript submitted November 29, 1955; made available for printing June 8, 1956.

RALPH KENNEDY is with the National Broadcasting Company, New York, N. Y.

The author is indebted to the American Telephone and Telegraph Company, which made available the color illustrations, Figs. 11, 12, and 13. Thanks are due to Howard Gronberg, National Broadcasting Company, for making available Fig. 14, and to the late R. B. Shelby and G. M. Nixon of the NBC Engineering Department for the encouragement they have given to the writing of this paper.





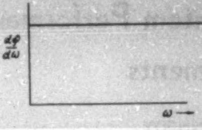


Fig. 1 (left). Ideal system time delay characteristic

Fig. 2 (center). Ideal system phase delay characteristic

Fig. 3 (right). Ideal system envelope delay characteristic

ference between these two frequencies is Δf. The phase detector must indicate the change in phase $\Delta\varphi$ between the two frequencies. Since twice the modulating frequency is Δf the absolute value of the two frequency components are the sum and difference of the suppressed carrier and the modulation frequency. As a result the test signal comprises two frequencies a constant Δf apart which scan the system phase characteristic as the carrier frequency is adjusted.

The remaining requirement for any type of phase measurement is a reference signal from the input with which to compare the output. For closed-circuit measurement, i.e., where the input and output terminals are available together, e.g., an amplifier, no problems are encountered on this score. However, for through system measurements where input and output might be miles apart, means must be provided to obtain a reference from the input to the output. Additionally, such systems are usually equipped with clamper amplifiers somewhere throughout their extent and need pulses for driving the clamps. These two requirements point to the use of pulses having a very stable frequency in conjunction with the two side-band components as a complete test signal. The two side-band components appear on an oscilloscope as shown in Fig. 5. Addition of the synchronizing signal results in a signal as shown in Fig. 6.

One device useful for making such measurement uses a 100-kc modulating signal or Δf of 200 kc and an adjustable signal generator having a frequency range from about 100 kc to 10 mc. Measurements can be made down to a frequency of about 350 kc. It should be realized that the synchronizing signal used as a reference, and which in this apparatus is 1/6 of the 100-kc signal, has higher order components with energies high enough to cause interference up to 200 or 300 kc. This acts as a lower limitation on the measurements.

The range between 50 and 350 kc needs additional investigation before thorough measurements are possible. Certainly the Δf could be reduced by using a slower rise-time synchronizing pulse so as to reduce the interference between them. However, this represents only a minor improvement. Some other approach must be found.

Such a test signal permits testing either closed or through circuits since the synchronizing pulses act as the reference signal. Closed-circuit measurements may also be made where $\Delta\varphi$ is kept constant and Δf is measured as has been pointed out by Kallman.[6]

Naturally, the question is asked: How does envelope delay affect a picture and to what extent? To answer this question it is necessary first to appreciate the effect the system having an ideal phase characteristic will produce on a signal when the system is ideally bandwidth limited. An ideal square wave will suffer deformation as predicted by Gibbs and shown in Fig. 7.

The effect produced on a square wave and additionally on an impulse has been determined by synthetic means. The synthesizer calibration is shown in Fig. 8(A) for an ideal bandwidth limited square wave and Fig. 9(A) for a $(\sin x)/x$ pulse. It is assumed that the envelope delay is constant between 50 kc and 4.2 mc. Figs. 8(B), 8(C), 9(B), and 9(C) show the effect for ± 0.05-μsec change while Figs. 10(A) and 10(B) are for ± 0.10-μsec variation. The overshoot for ±0.05-μsec variation is 14.6% (per cent) versus the 8.9% overshoot for the ideal case. This is evident as a leading or lagging black bar on the edges of any sharp black to white transitions in the picture. Tests show that 1% brightness variations are just visible; hence, nothing greater than 8.9% should be tolerated. In a color signal the same situation exists as for monochrome with the added difficulty of color being shifted with respect to outline detail. As an example, the red color of lips may be shifted with respect to monochrome lips appearing as though the lipstick had been displaced with respect to the outline of the lips. Such distortions are for envelope delay variations occurring in a uniform manner over a large portion of the frequency range.

On the other hand, abrupt delay variations occurring over rather small frequency increments in the upper portion of the video band can cause shifts in phase between various color components creating hue shift.

Differential Phase Distortion

Item 2 relates to the maintenance of close phase relationships between the various color components and between these components and burst. Departures from ideal are called differential phase distortion. In so far as the broadcaster is concerned, its cause may be summed up as due primarily to tube overloading though envelope delay distortions can also have an effect. This is particularly true for feedback circuits or circuits having slight regeneration at high frequencies.

Prior to color transmission it was common practice to give no consideration to differential phase distortion either

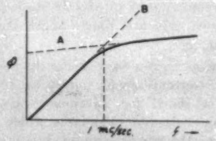


Fig. 4. Television transmitter side-band filter phase delay characteristic

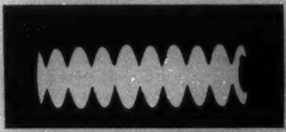


Fig. 5. Output of doubly balanced modulator

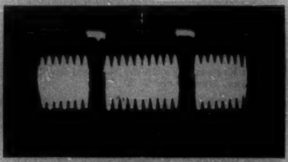


Fig. 6. Signal of Fig. 5 after inclusion of synchronous pulses

in audio or monochrome video amplifiers. It was customary to design feedback video amplifiers with 15,000 ohms in the plate circuit. The frequency response without feedback was flat to about 100 kc. The introduction of 30 to 35 db (decibels) of feedback resulted in a frequency response which was flat to 8 mc. However, the resultant improvement in the distortion applied only to the frequency band up to 100 kc. Above that frequency the amount of voltage available for feedback was decreasing and the distortion was increasing. However, the distortion did not produce too serious trouble, since the energy contained in the video signal decreased with frequency.

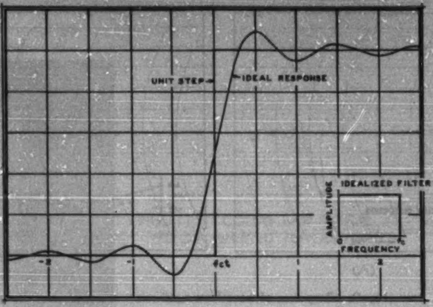


Fig. 7. Response of an ideal band-width limited system to a unit step

With the advent of color television this situation was altered when the amplifier was subjected to high-energy high-frequency components.[7] Inspection showed that a 3.579-mc test signal which is a sinusoid was warped into a sawtooth whose shape changes greatly for small changes in frequency. Fourier analysis of the wave forms showed that both sine and cosine terms were present in the series and that phase shift was introduced between the various components. It was found that the amount of phase shift was a function of the signal amplitude. Tests were devised and measurements made which showed that as much as 14 deg of differential phase shift was occurring in one amplifier. In some instances it was necessary to have six such amplifiers in a circuit.

The effect observed on a color picture which suffers differential phase distortion is one of hue shift as a function of signal amplitude.[8] In the case of saturated yellow, which is transmitted at the highest level of any component in the video signal, the color component was shifted so far in phase that it appeared green in the picture. See Figs. 11, 12, and 13.

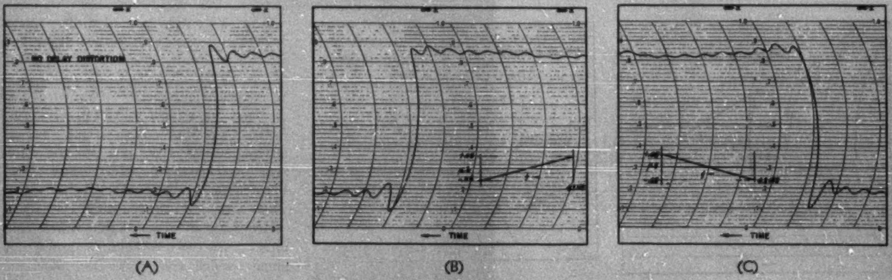
As one would expect, the low-amplitude components, e.g., suturated blue, magenta, and red, suffered least from this form of distortion while the effect of the distortion became more pronounced for higher amplitude components.

The measurement of differential phase distortion requires a special signal. Two methods are in current use.[9] If a Radio Corporation of America color signal analyzer is used the special signal is a 15.75-kc stair step having the bottom step at black and the top step at white, modulated with 3.579-mc sine wave having a level of about 20% of peak-to-peak video. Blanking and synchronization are added to this signal to create a conventional television signal. This signal is fed through the system under test and the differential phase shift measured by means of a color signal analyzer and an oscilloscope. The measurement shows the amount of shift in phase through the system for the small high-frequency sine wave signal at different stated levels of the low-frequency signal on which it is super-imposed. Shifts in phase of less than 1 deg are readily detected.

I AND Q VECTORS

The third type of phase distortion that must be controlled is the phase relationship which exists between the I and Q vectors. This distortion occurs in an improperly adjusted colorplexer. Like differential phase distortion, it makes itself evident in the picture as a shift in hue.

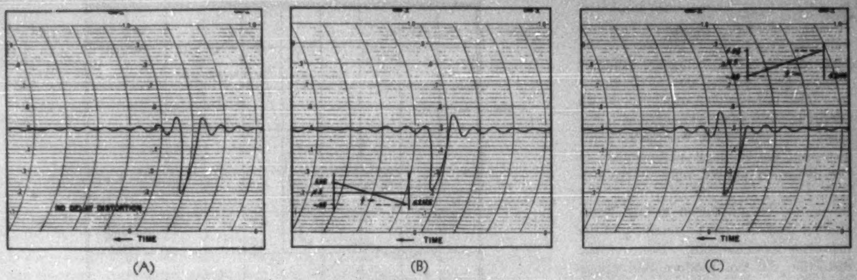
The Hazeltine Vectorscope is a display device intended to show the phase and amplitude of the chroma components and reference burst as well as that of the I and Q vectors. The standard color bar signal at the output of the colorplexer will present on the face of the Vectorscope the phase relationships that are produced by the encoder.



(A) (B) (C)

Fig. 8. Response of wave-form synthesizer adjusted for conditions of Fig. 7

A—No distortion
B,C—Distortion caused by indicated system envelope delay characteristics



(A) (B) (C)

Fig. 9. Response of wave-form synthesizer adjusted for unit impulse and ideal bandwidth limited system

A—No distortion B,C—Distortion caused by indicated system envelope delay characteristics

Influence of Amplitude on the Performance of the Color System

In speaking of amplitude effects on a color reproduction, two entirely different phenomena must be kept in mind. First, there is the usual amplitude versus frequency characteristic generally referred to as "frequency response." Second, there is what is known as differential gain. This gain was called compression in monochrome television. Differential gain more rigorously stated is the difference in gain of the system for a small high-frequency sine wave at various levels of a low-frequency signal on which it is superimposed.

Frequency Response

Originally it was thought that 1% allowable difference between the burst and horizontal line frequencies would be necessary, and with ±3% maximum allowable deviation in the region between them would be adequate for ten amplifiers. Experience has shown that this tolerance need not be this stringent.

The response variation of a system causes two effects to be observed in a color picture. Manifestly, a rolling-off of the high-frequency response curve starting at too low a frequency and changing at too rapid a rate will cause the picture to appear soft in monochrome and lack detail. In color the chromanence suffers desaturation as well.

From the tolerances which have been mentioned, the question undoubtedly arises as to how such measurements are made. The oscilloscope serves a very useful purpose when a 10-centimeter trace is used. One millimeter is equal to 1% and a centimeter scale permits checking system performance to the desired tolerance. The only problem is to calibrate the oscilloscope. However, there still remains the problem of measuring individual components of a system, e. g. amplifiers, equalizers, cables, and related apparatus.

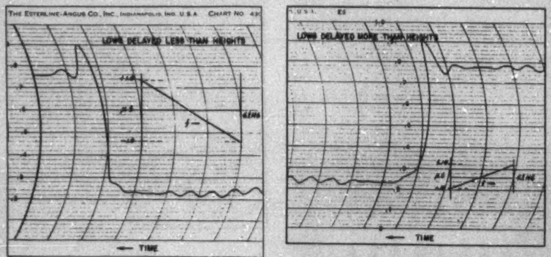
For a number of years the Bell System has used a device developed by the Bell Telephone Laboratories for making transmission measurements on video circuits. This is known as the type *70A* transmission measuring set. It is a thermocouple-operated power measuring meter having an impedance of 75 ±1 ohms. The meter is calibrated in decibels related to 1 milliwatt (dbm). The range is from −10 dbm to +3 dbm with 0 dbm equal to 1 milliwatt into 75 ohms. The scale length between +2 and +3 dbm is about 0.75 inch so that the 0.1-dbm scale designations are 0.075 inch apart. With such a scale, 0.05 dbm is easily read while values of 0.02 dbm are fairly accurately estimated. The meter has an accuracy of ±1%. The error of the meter plus meter reading error is about 1.3% in terms of absolute power; however, by using two meters for indicating relative input and output level, the measurement may be made to within less than a per cent.

Such meters not only permit making measurements on amplifiers, equalizer, cables, etc., but may be used as a standard for the calibration of oscilloscopes as well.

Differential Gain Distortion

The differential gain distortion is an indication of the compression characteristic of a circuit. The effective gamma range of the system is reduced. In a monochrome picture this form of distortion may make itself evident in one of two forms. The gamma range may be restricted in either the black or the white region. This results in a picture whose black areas tend toward gray or whose white areas tend toward gray.

A color picture may again suffer from either of the two forms of degradation. A circuit which compresses the black or restricts the gamma range in the dark portions of a monochrome picture will cause desaturation of purples and blues in a color picture. The opposite condi-



(A) (B)

Fig. 10. Distortion caused to wave form of Fig. 7 by indicated system envelope delay characteristics

Vol. 7
Pt. 1
1956

FIG. 11 — COLOR BAR SIGNAL — YELLOW SHIFTED TOWARD RED

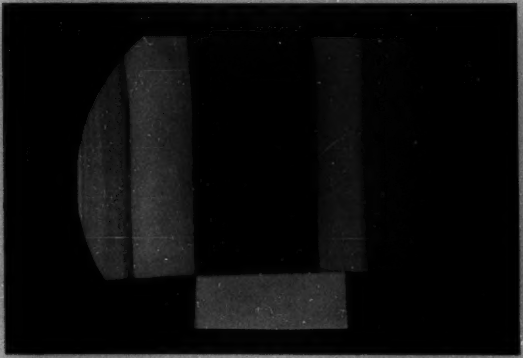


FIG. 12 — COLOR BAR SIGNAL — NORMAL PHASE

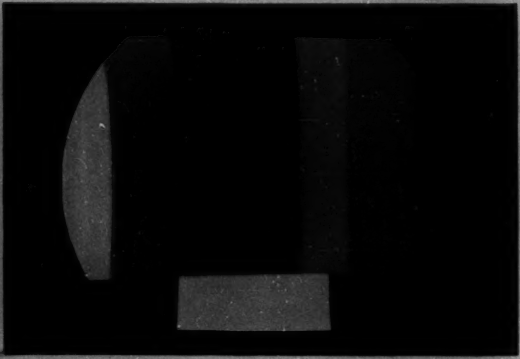


FIG. 13 — COLOR BAR SIGNAL — YELLOW SHIFTED TOWARD GREEN

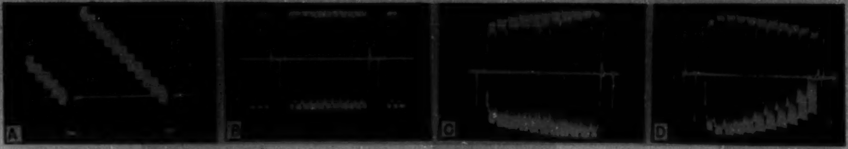


Fig. 14. Stair-step signal as originated, before (A) and after (B) going through high-pass filter; also, with 17% white compression (C) and with 8% white compression and 25% black compression (D)

tion wherein the monochrome whites are compressed results in the same washed-out appearance in the yellow areas of a color picture.

The cause of differential gain distortion is due to operating a tube in the system in the "knee" or highly nonlinear portions of its E_g-I_p characteristic. Operating the tube in class A and utilizing feedback are effective means for reducing the differential gain distortion.

As the name implies, this form of distortion is a function of the over-all signal amplitude. The measurement of such deterioration in a monochrome signal is accomplished by using a stair-step signal. Each step of the output signal must have the same relative amplitude as the input signal. The test signal for color systems is the same as is used for differential phase distortion measurements. After passing through the system the signal is passed through a 1-mc high-pass filter. This signal, when viewed on an oscilloscope, should have the same amplitude for each step. See Fig. 14.

Transient Measurement

SQUARE-WAVE TESTING

The literature is filled with material concerning transient testing of systems. By and large, the major share of the data are directed toward the square-wave or unit-step-function type of test signal. This is doubtless due to the fact that Heaviside concerned himself with this approach and his work was the basis for most subsequent related activity.[9]

There are a number of points which must be appreciated, however, regarding transient testing. The most important is the shape of the test pulse which will reveal the greatest amount of data about a system. This, in turn, calls attention to the type of data which the system is required to handle in normal operation. In the present instance it is a color television signal. This signal is a time sequence of electric impulses created by the horizontal passage of a circular cross-sectional beam of electrons across the photo image of the object viewed by the camera. Since the beam has a finite diameter, changes of brightness, represented by vertical edges in the picture, do not cause instantaneous variations in signal output level. On the contrary the voltage output follows a sin^2 shape very closely.[10] It is apparent that a sharp instantaneous rise as described by Heaviside is not a pulse which is truly representative of the camera output. Neither is the raw output of the present-day video square-wave generators which generally have rise times of the order of 0.02 μsec a suitable test signal. Indeed, the square wave, if it is to be used, should have a predetermined rise-and-fall characteristic following a sin^2 law.

It is interesting to consider the spectrum of the ideal square wave and the sin^2 shaped square wave. The ideal square wave has components which are spaced the repetition frequency apart out to infinity. The sin^2 law causes the pulse spectrum to be 6 db down at a frequency equal to half the reciprocal of the rise time between zero and 100%. At twice this frequency the spectrum drops to zero and remains about 30 db down for all components above that frequency. Clearly, the shaped square wave has a spectrum which is a much more realistic test for the system.

As noted, the rise-and-fall time determines the spectral distribution. To test the system properly it appears that two different pulse shapes would be desirable. When horizontal line frequency used as the repetition rate, there should be one pulse having a spectrum zero at 4.0 mc and the other at 8.0 mc. This provides signals capable of evaluating the system in the middle and upper portions of the frequency band.

The measurement of transient response using square waves is more generic than specific. The display on the oscilloscope requires long and painstaking effort to acquire analytical data about a system.[11] Generally, there are four questions which can be asked about the transient characteristics of systems: 1. If the wave form of several links be known, what is the wave form when they are cascaded? 2. Given the wave form for a circuit composed of a number of identical links cascaded, what is the wave form of one of the links? 3. Given the wave form for a circuit and a part, what is the wave form of the remainder? 4. Given the response to one wave form, what is the response to

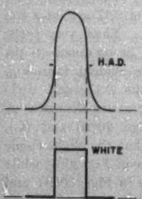


Fig. 15. Wave form produced by scanning a single picture element

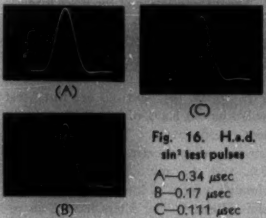

(A)
(C)
(B)

Fig. 16. H.a.d. sin^2 test pulses
A—0.34 μsec
B—0.17 μsec
C—0.111 μsec

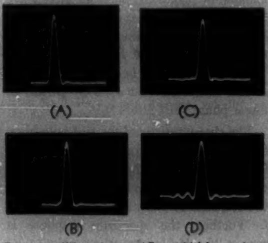

(A)
(C)
(B)
(D)

Fig. 17. Wave forms of Fig. 16(A) at various filter outputs
A—6.0-mc filter
B—4.0-mc filter
C—2.3-mc filter
D—1.67-mc filter

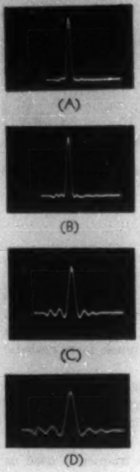


Fig. 18. Wave forms of Fig. 16(B) at various filter outputs

A—6-mc filter
B—4-mc filter
C—2.3-mc filter
D—1.67-mc filter

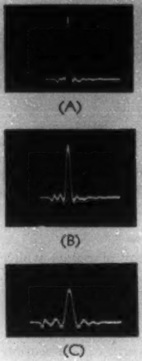


Fig. 19. Wave forms of Fig. 16(C) at various filter outputs

A—6-mc filter
B—4-mc filter
C—2.3-mc filter

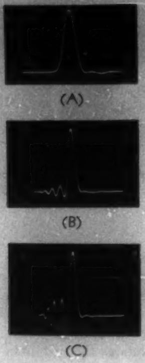


Fig. 20. Wave forms of Figs. 16(A), (B), and (C) at output of phase corrector

another? The answers to questions of this type are important and yet require considerable effort. The usual square-wave tests involve an attempt to measure rise time which gives a crude indication of bandwidth. The degree of nonsymmetry about the transition which is produced by phase distortion is at best only a rough approximation.

IMPULSE TESTING

The use of square waves in conjunction with wave-form correctors provides a fairly satisfactory method of equalizing a circuit for transient response. However, there appears to be considerable advantage to be gained by using the shaped impulse. Ideally it should be a $(\sin x)/x$ function. However, it is impossible to construct a shaping filter capable of producing such an ideal wave form. A $\sin^2 x$ shaping filter may be built which causes the pulse to conform to the mathematical ideal within 1%.[12] This pulse may then be shaped roughly to approximate the $(\sin x)/x$ by means of linear phase shift filters.[13,14]

The characteristic of the $\sin^2$ pulse may be related to the spectrum response as follows: When the half-amplitude duration (h.a.d.) of the pulse is equal to $1/2f$, the response is down 6 db at a frequency f. Further, the spectrum response is zero at $2f$. If f is 4 mc (the nominal television bandwidth), the h.a.d. is 0.125 μsec. The spectral response of the pulse is 6 db down at 4.0 mc and is zero at 8 mc. Further, the diameter of the smallest black or white element which can be sent over the system is equal to the h.a.d. of the pulse. See Fig. 15. Ideal bandwidth limiting causes a 13% overshoot versus 8.9% for a step function. Further, the area under the pulse remains constant for all degrees of band limiting.

The generation and display of such a pulse requires special apparatus. The generator is driven by horizontal pulses which are shaped and drive a blocking oscillator. The output is a very short duration impulse, which is formed into a $\sin^2$ pulse in a shaping filter. One of the most desirable features of this test means is the ease of duplication of the test pulse. The shaping filter which consists of five capacitors and four coils controls the entire shaping operation. So long as these elements are measured to within 1%, the pulse output deviates from the ideal by less than 1%. Clearly no problem exists in building generators having the same output wave form.

The Tektronix *524* has been used as a display means and thus far has been satisfactory for viewing the pulse out of the shaping filter. A better oscilloscope, however, is required to view the impulse into the shaping filter since it has an h.a.d. of about one fifth that of the output pulse.

The interpretation of the wave forms is quite easy. Figs. 16–20 show $\sin^2$ pulses of various h.a.d. together with the effects produced, by various degrees of bandwidth limitation and phase nonlinearities. Figs. 16(A), (B), and (C) are three test pulses having h.a.d. of 0.34, 0.17, and 0.11 μsec, respectively. These pulses have corresponding spectra down 6 db at 1.5, 3.0, and 4.5 mc. They were passed through a series of low-pass linear-phase-shift filters.[14] The filters have 3-db rejection frequencies at 6.0, 4.0, 2.3, and 1.67 mc. Figs. 17(A), (B), (C), and (D) show how the filters effect the 0.34-μsec pulse. Observe that Figs. 17(C) and (D) show a decreasing pulse height and increasing width. Also note the symmetry in Fig. 17(D) as the pulse begins to assume a $(\sin x)/x$ shape. Since the pulse contains essentially no energy above 3.0 mc, the 6.0- and 4.0-mc filters have no effect on the pulse height or duration. This is observed in Figs. 17(A) and (B).

Figs. 18(A), (B), (C), and (D) show similar bandwidth reductions on the 0.17-μsec pulse. Here reduced pulse height and increased width are very apparent.

Figs. 19(A), (B), and (C) illustrate the effect that the 6.0-, 4.0-, and 2.3-mc filters produce on the 0.11-μsec pulse. It is quite evident that the phase characteristic of the 6.0- and 4.0-mc filters is not so good as the 2.3- and 1.67-mc types as indicated by the lack of symmetry of the ripples.

Figs. 20(A), (B), and (C) show the deformation of the pulse caused by a lossless nonlinear network.[15] The particular network used was intended to produce the envelope delay characteristic for the color-system television transmitters. This requires that the envelope delay remain constant from 50 kc to 3 mc and then decrease linearly to 4.18 mc resulting in a value of 0.17 μsec at a frequency of 3.579 mc. Thus, very little, if any, effect should be produced on the 0.34-μsec pulse as is evident from Fig. 20(A). However, the 0.17- and 0.11-μsec pulses suffer increasing distortion as is evident from Figs. 20(B) and (C). Furthermore, note that the slight ripple follows the main pulse in Fig. 20(A) while it precedes the pulse in Figs. 20(B) and (C). If the network has an increasing time of transmission with frequency (the opposite of the network used), the location of the ripple is reversed. This reveals the type of phase shift present in a system.

In Europe where sin² testing techniques have reached a high degree of precision, a system test signal is in common use which consists of four different wave forms. The low-frequency portion of the system is adjusted by use of a 50-cycle square wave. This corresponds to the 60-cycle square wave in the 525-line system.

In addition, a conventional television signal is generated which consists (for the 525-line system) of a 30-μsec shaped bar having amplitude of standard black-to-white video. On the same horizontal line and preceding the bar is inserted either of two sin² pulses. One pulse has an h.a.d. of 0.125 μsec and the other is 0.25 μsec.

This combined test signal provides a ready means of ascertaining the streaking or smearing potential of a circuit which occurs in the frequency region from 15 kc to a few hundred kc. This is indicated by the 30-μsec bar. By having a rise and fall the same as the 0.25-μsec h.a.d. sin² pulse the bar will contain components 6 db down at 2.0 mc and will have no energy at 4.0 mc. The high-frequency characteristic of the system causes little, if any, distortion to the bar.

The 0.25-μsec sin² pulse gives sensitive indications to transient distortions in the middle of the video band while the 0.125-μsec pulse is sensitive to the variations around 4.0 mc.

The British Broadcasting Corporation and the Post Office Research Department have been using this technique for a number of years and some indication of the sensitivity of the test may be obtained when it is realized that it is possible easily to detect certain forms of distortion with 0.2-db attenuation distortion and 0.02-μsec delay distortion in a 3-mc system.

In addition the methods of using the test are divided into two categories.[16] There is a procedure known as the acceptance test which is used for the complete initial adjustment of a studio or circuit and there is the routine-testing procedure which is used as a quick maintenance test or for equalizing temporary circuits to remote pickup locations.

In addition to the acceptance test procedure there has been developed a mathematical procedure using time series which permits making calculations of and obtaining answers to the aforesaid four questions, 1–4.[17] Further, a system rating factor is proposed to permit comparison of wave-form responses of different systems.

The routine-testing procedure is greatly enhanced by the wave-form corrector which permits a rapid and accurate means of equalizing a remote program circuit.[18]

Conclusions

As is true in any type of communication system when it is necessary to send more information either a wider frequency band is necessary or more stringent tolerances must be placed on the various performance requirements of the existing circuit. The latter requirement obtains in the case of the color system and is brought about by the chroma being transmitted in the high-frequency portion of the video band where small monochrome energies are to be found.

While a monochrome signal suffers less distortion when sent through a system intended for color, it certainly does not require such a transmission path to result in a perfectly usable picture.

Envelope delay, while critical at lower frequencies in monochrome, probably does not need to be so precisely controlled to 4.2 mc as it does for color. Differential phase distortion certainly is color signal degradation as is the **I** and **Q** vector relationship.

The frequency response of a monochrome system was generally maintained fairly constant so as to assure a reasonable phase characteristic rather than to avoid a loss in picture detail. Certainly no ±3% response was required in the video band to produce a good picture.

Kinescope recording placed about the most exacting demands on the gamma characteristic of a monochrome system. Gamma correctors were available to produce what was desired rather than having to redesign the whole system.

Transient studies of monochrome systems were made by different investigators but generally such studies were not directed to so precise a solution of the transient problem. It appears that one of the largest areas for future transient studies is the per-cent rating factor method. This is particularly true when one realizes that once a number has been accepted as representing "broadcast quality," any amplifier, studio, transmitter, or circuit may be judged on a purely transient basis and immediately one knows the acceptability of such equipment.

References

1. Rules Governing Broadcast Services. Federal Communications Commission, Washington, D. C., part 3.687.

2. Communication Networks, Vol. 2 (book), **E. A. Guillemin.** John Wiley & Sons, Inc., New York, N. Y., 1935, p. 490.

3. Measurement of Phase Distortion, **H. Nyquist, S. Brand.** *Bell System Technical Journal*, New York, N. Y., vol. IX, July 1930, pp. 522–43.

4. Influence of Echoes on Television Transmission, **P. Mertz.** *Journal*, Society of Motion Picture and Television Engineers, New York, N. Y., vol. 60, May 1953, pp. 572–96.

5. Color Video Envelope Delay Measurement, **R. C. Kennedy, H. French.** *Electronics*, New York, N. Y., vol. 28, no. 6, June 1955, pp. 144–48.

6. Equalized Delay Lines, **H. E. Kallman.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 34, Sept. 1946, pp. 646–57.

7. Color-Video Transmission Over Intercity Television Networks, **G. J. Harms.** *Electrical Engineering*, vol. 74, Aug. 1955, pp. 667–70.

8. Television Signal Analysis—Long Lines Department (book). American Telephone and Telegraph Company, New York, N. Y., May 1955, p. 47.

9. Electromagnetic Theory, Vol. 2 (book), **Oliver Heaviside.** Dover Publications, Inc., New York, N. Y., 1950.

10. High Power Television Transmitter Technique with Particular Reference to the Transmitter at Holme Moss, **V. J. Cooper.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. III-A, 1952, p. 231.

11. Analysis Synthesis and Evaluation of the Transient Response of Television Apparatus, **A. V. Bedford, G. L. Fredenndall.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 30, Oct. 1942, p. 440.

12. A Testing Pulse for Television Links, **I. F. Macdairmid.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. III-A, 1952, p. 436.

13. Ideal Wave Filters, **H. W. Bode, R. L. Dietzold.** *Bell System Technical Journal*, New York, N. Y., vol. 14, Apr. 1935, p. 215.

14. Linear Phase Shift Video Filters, **G. L. Fredendall, R. C. Kennedy.** *RCA Review*, Princeton, N. J., vol. 9, Sept. 1950, p. 418.

15. Delay Equalization in a Color Television, **G. L. Fredendall.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, Jan. 1954, p. 258.

16. Waveform Responses of Television Links, **N. W. Lewis.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 101, pt. III, no. 72, July 1954.

17. Waveform Computations by the Time-Series Method. *Ibid.*, vol. 99, pt. III-A, 1952, p. 294.

18. A Variable Time-Equalizer for Video-Frequency Waveform Correction, **J. M. Linke.** *Ibid.*, p. 427.

The Simple Reactor Circuit; Its Operation and Mode Transition

J. F. RINGELMAN
ASSOCIATE MEMBER AIEE

A. L. FENAROLI
ASSOCIATE MEMBER AIEE

METHODS for designing simple saturable-reactor magnetic amplifiers and predicting their performance on the basis of observed phenomena have been known for some time. However, the literature in the field is still lacking a complete analysis of saturable-reactor circuitry on the basis of the mechanics of operation. Ramey[1] approached the problem of saturable-reactor analysis from a novel point of view. The theory and relationships developed in his paper established a firm foundation for the understanding of the amplifier operation. This paper utilizes the volt-second approach developed by Ramey to analyze completely the simple series saturable-reactor circuit in both its transient and steady states.

The simple reactor circuit is a complex device to analyze. An analysis based upon nonideal flux current loops leads to involved mathematical expressions which tend to obscure the mechanics of operation. Similarly, an analysis considering ideal loops will also obscure the concept of operation if sine wave excitation and constant d-c control are used for forcing functions. This follows from the fact that the resulting circuit equations do not lend themselves to explicit solutions but must be solved graphically. Since it is the purpose of this paper to describe accurately the mechanics of operation of the simple reactor circuit, two simplifying ideas are utilized. First, the analysis assumes an ideal flux current loop of zero width, an infinite slope in the unsaturated region, and a zero slope in the saturated region. Second, a square-wave supply voltage is used.

The use of an ideal loop is a valid assumption in the case of simple saturable-reactor circuits because core losses usually have only second-order effects in practical circuits and are superfluous to the mechanics of operation. If ideal loops are assumed, it follows that the reactors are sensitive to volt-second integrals only, regardless of the voltage wave shapes supplied by the sources. For purposes of analysis it is advantageous to select forcing functions which will yield explicit solutions as simply as possible. The solutions thereby derived can then be applied to saturable-reactor circuits utilizing any periodic a-c exciting voltage in the power frequency range. The combination of a square-wave exciting voltage and a constant d-c control voltage will yield explicit solutions. In addition, the relative magnitude of effective control voltage to a-c excitation voltage for this particular combination is clearly defined at all times. The latter point is extremely important in determining the mode in which the circuit is operating.

The actual circuit to be analyzed is a series-connected simple reactor circuit controlled by a constant-amplitude direct voltage and supplied by an alternating voltage of square wave shape. The analysis is restricted to resistive loads and resistive control circuit impedances. The circuit operation is described for all ratios of control circuit resistance to load resistance and control voltage to supply voltage.

The analysis as presented is based upon period equations. Each half-cycle of the supply voltage is divided into two periods, a conducting and a nonconducting period. For each period, enough is known to allow for the derivation of equations for load current, control current, and voltages across the reactors. From these equations, expressions describing the transient and steady-state operation of the circuit are derived.

With the aid of the one assumption, i.e., an ideal flux current loop, and the use of a square-wave supply voltage, several important relations describing steady-state and transient operation of the circuit are derived. The relative unimportance of the supply voltage wave shape employed to the actual mechanics of operation is emphasized by some of the derived equations which are identical with those derived by Ramey, utilizing completely different forcing functions. The circuit operation in each half-cycle is shown to depend upon the operation in the previous half-cycle. The output current is shown to be transient only in respect to firing angle. Two modes of operation are shown to exist which are fundamentally different in nature. Finally, the transition from mode to mode is shown to be smooth and to depend upon circuit parameters and relative magnitudes of the source voltages present in the circuit.

List of Symbols

e_{ac} = instantaneous a-c supply voltage
e_L = instantaneous load voltage
E_{ac} = instantaneous square-wave supply voltage
$E_{L,1,avg}$ = average load voltage for first half-cycle
$E_{L,n,avg}$ = average load voltage for nth half-cycle
$E_{L,ss}$ = average steady-state load voltage
E_m = maximum value of sine wave
E_c = d-c control voltage
i_L = instantaneous load current
i_c = instantaneous control current
I_m = instantaneous magnitude of load current in conducting period (square wave)
$I_{L,n,avg}$ = average load current in nth half-cycle
I_{ss} = average steady-state load current
$I_{c,avg}$ = average control current
R_c = lumped control circuit resistance
R_L = lumped load circuit resistance
t_n = firing time in the nth half-cycle
t_{ss} = steady-state firing time
T = time duration of half-cycle
N = turns ratio (control turns)/(load turns)
n = integer number of half-cycles greater than zero

$$K=N^2\frac{R_L}{R_c}$$

$$A=\frac{E_c/N}{E_{ac}}$$

$$B=\frac{1-AK}{1+AK}$$

Analysis

The circuit under consideration is shown in Fig. 1. The control voltage E_c is a constant-amplitude direct voltage and e_{ac} is the instantaneous magnitude of, for the present, any periodic alternating voltage. All the resistances of the control and load circuits are lumped into R_c and R_L respectively. Fig. 2 illustrates the flux current loops of reactors A and B.

The general equations for the load and control circuits may be written as

$$e_{ac}=e_a+e_b+i_LR_L \tag{1}$$

$$E_c=Ne_a-Ne_b+i_cR_c \tag{2}$$

where N = (control turns)/(load turns). It is now advantageous to define the conducting and non-conducting periods of a half-cycle which depend upon the relative flux conditions of the two reactors. A

Paper 56-750, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 27, 1956; made available for printing May 15, 1956.

J. F. RINGELMAN and A. L. FENAROLI are with Westinghouse Electric Corporation, Baltimore, Md.

Vol.
Pt. 1
1956

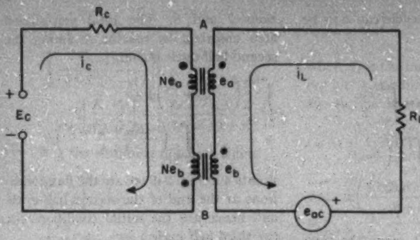


Fig. 1. Series-connected simple saturable reactor circuit

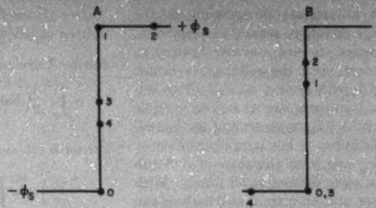


Fig. 2. Idealized flux current loops for reactors A and B

nonconducting period exists when both reactors are simultaneously unsaturated. Since ideal flux current loops have been assumed and exciting currents do not exist, control and load currents will be zero. Thus, during a period of nonconduction the general equations 1 and 2 may be reduced to the form

$$e_{ac}=e_a+e_b \qquad (3)$$

$$E_c=Ne_a-Ne_b \qquad (4)$$

Now, if one reactor saturates while the other remains unsaturated, the voltage across the saturated reactor will drop to zero. The conduction period begins at the instant of saturation of one of the reactors, allowing the other reactor to act as a transformer, and continues until both reactors are again unsaturated. Load and control currents will flow during the conduction period and, since the unsaturated reactor must still carry zero net ampere turns, $N_c i_c = N_L i_L$ or

$$i_c=\frac{i_L}{N}$$

During a conduction period when reactor *A* is saturated and reactor *B* is unsaturated, equations 1 and 2 may be reduced to

$$e_{ac}=e_b+i_L R_L \qquad (5)$$

$$E_c=-Ne_b+i_c R_c \qquad (6)$$

By making use of the relationship $i_c = i_L/N$, equations 5 and 6 may be solved for the currents

$$i_L=\frac{e_{ac}+\dfrac{E_c}{N}}{\dfrac{R_c}{N^2}+R_L} \qquad (7)$$

$$i_c=\frac{Ne_{ac}+E_c}{R_c+N^2R_L} \qquad (8)$$

Equations 7 and 8 indicate the current wave forms that can be expected for relative magnitudes of control and supply voltages. If e_{ac} is a sinusoid and E_c a constant-amplitude direct voltage, it is apparent that the amount of roundness or squareness of the current pulses is dependent upon the relative magnitudes of E_c/N and e_{ac}. In addition, the polarities of the control windings, during the conduction portion of each half-cycle, are such as to transform the supply voltage into the control circuit always in the same direction. The alternating current is thus inverted on alternate half-cycles, and a double frequency alternating current is introduced into the control circuit. This is sometimes referred to as a second harmonic. The relative content of the double frequency alternating current in the control circuit is dependent upon the relative magnitude of E_c and Ne_{ac}. Of importance is the fundamental relationship that in the conducting period the instantaneous control ampere-turns applied to the unsaturated reactor must equal the instantaneous load ampere-turns. Since, during the nonconducting period, the control and load currents are zero, the average value of control ampere-turns equals the average value of the load circuit ampere-turns for the entire half-cycle. It is not apparent, from equation 8, that the average value of the steady-state control current for the entire half-cycle is E_c/R_c. However, this is true since direct current cannot be transformed; this will be discussed later.

Thus far, the load and control currents have been described in the conducting and nonconducting periods. To define completely the circuit, it remains to consider it in its transient state. Before proceeding, it is desirable to establish initial conditions. The supply voltage is chosen to be of square wave form, and its magnitude at any instant described by E_{ac}. The flux condition for each reactor is assumed to be initially at negative saturation, point zero in Fig. 2. Referring to Fig. 3, a transient is introduced at $t=0$, at which time the supply voltage is zero and increasing positively, by the application of a control voltage E_c. Each reactor is designed to absorb one-half the volt-second area available from the supply. Solving equations 3 and 4 for e_a and e_b in the nonconducting period, reactor *A* has applied to it the voltage 1/2 $(E_{ac}+E_c/N)$ and reactor *B*, 1/2 $(E_{ac}-E_c/N)$. It is readily apparent that any voltage E_c, however small, will cause reactor *A* to saturate at some time before the end of the first positive half-cycle. The flux excursion of reactor *B*, however, is not so apparent. It is possible for reac-

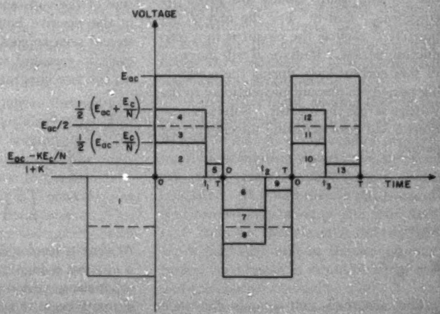


Fig. 3. Typical reactor voltages during a transient in mode 1 with K>1

tor B to saturate immediately in the negative direction, remain at point zero, or proceed toward positive saturation, dependent upon the relative magnitude of E_c/N with respect to E_{ac}. Two different transients are therefore possible with $E_c/N = E_{ac}$ representing the boundary condition. For a supply voltage of square wave form the ratio of E_c/N to E_{ac} is clearly defined at all times. With this fact utilized a transient analysis can be developed for each case. The first case to be considered is described by $E_c/N < E_{ac}$.

MODE 1: $E_c/N < E_{ac}$

If E_c/N is less than E_{ac}, reactor A will proceed toward positive saturation and saturate first at some time before the end of the first positive half-cycle. Reactor B, in the meantime, will reach some point on the unsaturated portion of its flux current loop. These conditions are illustrated by points 1 in Fig. 2.

Fig. 3 shows the reactor voltages during a transient with the circuit in mode 1 operation. At time $t = t_1$ reactor A saturates. Since each reactor is capable of absorbing one-half the volt-second area available from the supply voltage, the time t_1 is defined by

$$\int_0^T \frac{E_{ac}}{2}\,dt = \int_0^{t_1} \frac{1}{2}\left(E_{ac} + \frac{E_c}{N}\right) dt$$

where T = time duration of a half-cycle. From $t = 0$ to $t = t_1$ reactor B proceeds toward positive saturation by absorbing the volt-second integral

$$\int_0^{t_1} \frac{1}{2}\left(E_{ac} - \frac{E_c}{N}\right) dt$$

At time $t = t_1$ reactor A saturates and is effectively eliminated from the load and control circuits. Reactor B is still capable of absorbing volt-seconds and hence operates as a transformer. Control current and load current begin to flow immediately and their values are given by equations 7 and 8. For the rest of the positive half-cycle the circuit is in a conducting period. Reactor B will move on its flux current loop, during the conducting period, an amount proportional to the voltage across the reactor. Solving equations 5 and 6 for e_b during the conducting period

$$e_b = \frac{E_{ac} - KE_c/N}{1+K}$$

where

$$K = \frac{N^2 R_L}{R_c}$$

The absorbed volt-seconds for each reactor for the first half-cycle can now be written as follows:

reactor A volt-seconds

$$= \int_0^{t_1} \frac{1}{2}\left(E_{ac} + \frac{E_c}{N}\right) dt = \int_0^T \frac{E_{ac}}{2}\,dt \quad (9)$$

$$\text{reactor } B \text{ volt-seconds} = \int_0^{t_1} \frac{1}{2}\left(E_{ac} - \frac{E_c}{N}\right) dt + \int_0^{t_1} \left(\frac{E_{ac} - KE_c/N}{1+K}\right) dt \quad (10)$$

Solving equation 9 for t_1 gives the time for the first firing angle

$$t_1 = \frac{T}{1 + \dfrac{E_c}{NE_{ac}}} \quad (11)$$

At the end of the first half-cycle the reactor fluxes are at position 2 in Fig. 2 and currents have flowed during the conducting period, as given by equations 7 and 8. The initial conditions for the second or negative half-cycle are therefore established. In a similar manner, the initial conditions in each half-cycle are established by the previous half-cycle. Hence for the purpose of writing equations involving volt-second areas each half-cycle can be treated individually as starting at time $t = 0$ with its associated initial conditions.

For the second half-cycle the polarity of the supply voltage changes and both reactors are again in a position to change flux. Solving for the voltage applied to the reactor B from equations 3 and 4, e_b is found to be $-E_{ac}/2 - E_c/2N$. This voltage is equal in magnitude to the voltage applied to reactor A in the nonconducting portion of the previous half-cycle. Due to the initial conditions for the second half-cycle, reactor B requires less flux change to saturate than did reactor A in the first half-cycle. Thus reactor B will saturate at a time t_2 which will be less than t_1. The time that reactor B saturates is determined by equating the sum of the volt-second areas it absorbed in the positive half-cycle and the volt-second area applied in the negative half-cycle to zero and solving for t_2. Equating the two integrals yields

$$-\int_0^{t_2} \frac{1}{2}\left(-E_{ac} - \frac{E_c}{N}\right) dt = \int_0^{t_1} \frac{1}{2}\left(E_{ac} - \frac{E_c}{N}\right) dt + \int_{t_1}^T \left(\frac{E_{ac} - KE_c/N}{1+K}\right) dt \quad (12)$$

At time t_2 reactor B saturates, reactor A is at point 3, Fig. 2, and the circuit is in a conducting period. Load and control currents begin to flow according to equations 7 and 8. The total volt-second area absorbed by reactor A during the second half-cycle is then

$$\int_0^{t_2} \frac{1}{2}\left(-E_{ac} + \frac{E_c}{N}\right) dt + \int_{t_2}^T \left(\frac{-E_{ac} + KE_c/N}{1+K}\right) dt$$

Points 4 in Fig. 2 illustrate the flux conditions at the end of the second half-cycle and hence are the initial conditions for the third half-cycle.

For the third half-cycle reactor A will saturate at time t_3 when it absorbs a positive volt-second area equal to the negative volt-second area it absorbed in the second half-cycle. If the change is observed in the direction of flux change this can be written as

$$\int_0^{t_3} \frac{1}{2}\left(E_{ac} + \frac{E_c}{N}\right) dt = -\int_0^{t_2} \frac{1}{2}\left(-E_{ac} + \frac{E_c}{N}\right) dt - \int_{t_2}^T \left(\frac{-E_{ac} + KE_c/N}{1+K}\right) dt \quad (13)$$

If signs are cleared in equations 12 and 13, the expressions are seen to be identical. Therefore, the general expression for the nth half-cycle firing time can be written as

$$\int_0^{t_n} \frac{1}{2}\left(E_{ac} + \frac{E_c}{N}\right) dt = \int_0^{t_{n-1}} \frac{1}{2}\left(E_{ac} - \frac{E_c}{N}\right) dt + \int_{t_{n-1}}^T \left(\frac{E_{ac} - KE_c/N}{1+K}\right) dt$$

or

$$\int_0^{t_n} \frac{1}{2} E_{ac}(1+A)\,dt = \int_0^{t_{n-1}} \frac{1}{2} E_{ac}(1-A)\,dt + \int_{t_{n-1}}^T E_{ac} \frac{1-AK}{1+K}\,dt \quad (14)$$

where $A = (E_c/N)/E_{ac}$

Solving equation 14 for t_n in terms of t_{n-1} gives

$$t_n = \frac{K-1}{K+1} t_{n-1} + 2\left[\frac{1-AK}{(1+K)(1+A)}\right] T \quad (15)$$

Solving equation 15 for t_2, t_3, t_n in terms of t_1 gives

$$t_2 = \frac{K-1}{K+1} t_1 + 2\,\frac{1-AK}{(1+K)(1+A)}\,T$$

$$t_3 = \left[\frac{K-1}{K+1}\right]^2 t_1 + 2\left[\frac{K-1}{K+1}\right]^1 \times \left[\frac{1-AK}{(1+K)(1+A)}\right] T + 2\left[\frac{1-AK}{(1+K)(1+A)}\right] T$$

$$t_n=\left[\frac{K-1}{K+1}\right]^{n-1}t_1+2\left[\frac{K-1}{K+1}\right]^{n-3}\times \left[\frac{1-AK}{(1+K)(1+A)}\right]T+\ldots+ 2\left[\frac{K-1}{K+1}\right]^{n-n}\left[\frac{1-AK}{(1+K)(1+A)}\right]T$$

or, using the summing notation sigma

$$t_n=\left[\frac{K-1}{K+1}\right]^{n-1}t_1+ \sum_{p=2}^{p=n}\left[\frac{K-1}{K+1}\right]^{n-p}2T\left[\frac{1-AK}{(1+K)(1+A)}\right]$$

The second term of the expression for t_n is a geometric progression of $n-1$ terms, where n is the number of half-cycles, of the form

$$\sum_{p=2}^{p=n}\alpha r^{n-p}$$

The sum of $n-1$ terms is

$$\frac{\alpha(1-r^{n-1})}{1-r}$$

where $\alpha=2T\left[\frac{1-AK}{(1+K)(1+A)}\right]$

and $r=\frac{K-1}{K+1}$

The general expression for t_n in terms of T, utilizing equation 11 for t_1, is

$$t_n=\frac{T}{1+A}\left[\frac{K-1}{K+1}\right]^{n-1}+ \frac{2T\left[\frac{1-AK}{(1+K)(1+A)}\right]\left[1-\left(\frac{K-1}{K+1}\right)^{n-1}\right]}{1-\frac{K-1}{K+1}} \qquad (16)$$

Once the transient has ended, the circuit is in steady state and the firing angle remains constant.

$$t_n=t_{n-1}=t_{ss}$$

Making this substitution in the general expression for t_n in equation 15 yields

$$t_{ss}=\frac{1-AK}{1+A}T \qquad (17)$$

As has been pointed out, the transient which ensues, upon the application of a control voltage, concerns only the firing angle. The average value of the currents flowing are, of course, dependent upon the firing angle; however, the currents that flow are known for the entire half-cycle and merely the firing time is necessary to solve for the average current. Therefore the average current in the nth half cycle is

$$I_{L,n,\mathrm{avg}}=I_M\left[\frac{T-t_n}{T}\right]=I_M\left[1-\frac{t_n}{T}\right] \qquad (18)$$

where I_m is the magnitude of the load current in a conducting period. The average current in steady state is

$$I_{ss}=I_M\left[1-\frac{t_{ss}}{T}\right], \qquad (19)$$

Dividing $I_{L,n,\mathrm{avg}}$ by I_{ss} yields

$$\frac{I_{L,n,\mathrm{avg}}}{I_s}=\frac{1-\frac{t_n}{T}}{1-\frac{t_{ss}}{T}}$$

Now, substituting for t_n and t_{ss} from equations 16 and 17 yields

$$\frac{I_{L,n,\mathrm{avg}}}{I_{ss}} =\frac{1-\frac{1}{T}\left\{\frac{T}{1+A}\left[\frac{K-1}{K+1}\right]^{n-1}+ 2T\frac{\left[\frac{1-AK}{(1+K)(1+A)}\right]\left[1-\left(\frac{K-1}{K+1}\right)^{n-1}\right]}{1-\frac{K-1}{K+1}}\right\}}{1-\frac{1}{T}\left(\frac{1-AK}{1+A}\right)t}$$

Solving for $I_{L,n,\mathrm{avg}}$ yields

$$I_{L,n,\mathrm{avg}}=I_{ss}\left[1-\frac{K}{K+1}\left(\frac{K-1}{K+1}\right)^{n-1}\right]$$

Since the average load current is equal to N times the average control current

$I_{ss}=N\frac{E_c}{R_c}$ and

$$I_{L,n,\mathrm{avg}}=\frac{NE_c}{R_c}\left[1-\frac{K}{K+1}\left(\frac{K-1}{K+1}\right)^{n-1}\right] \qquad (20)$$

The transient for the load current is completely described by equation 20. If $I_{L,n,\mathrm{avg}}$ is made equal to $1-1/e$ and the equation solved for $n/2$, the resulting equation is identical to that derived by Ramey. It is interesting that Ramey utilized completely different forcing functions, i.e., a sine wave supply and a rectified sine wave for d-c control.

It has become apparent through the development of the transient equation for load current that the transient proceeds according to relationships of the volt-second integrals applied to the two reactors. With the aid of Fig. 3 one may now more clearly visualize the transient. This figure shows the source voltage E_{ac} with respect to reactor voltages over three half-cycles. If the circuit is that of Fig. 1 and it is again assumed that the transient is initiated by the application of E_c at the beginning of the first positive half-cycle and that both reactors are at negative saturation, the transient will progress as follows. Each reactor is capable of absorbing a volt-second area equal to 1/2 $E_{ac}T$ or area 1. The voltage across reactor A, which will saturate in the first half-cycle, is 1/2 $(E_{ac}+E_c/N)$. This reactor will saturate at time t_1 when areas 2+3+4 equal area 1. Meanwhile reactor B, in proceeding from negative saturation, absorbs area 2. At time t_1 the reactor A saturates and absorbs no more volt-seconds for the remainder of the half-cycle. Reactor B, acting as a transformer, proceeds farther toward but does not reach positive saturation by absorbing area 5.

In the first half-cycle, then, the area absorbed by reactor A is area (2+3+4); and that absorbed by reactor B is area (2+5). If K is greater than 1, area (2+5)<area (2+3+4).

In the second half-cycle, reactor B now has applied to it the voltage minus 1/2 $(E_{ac}+E_c/N)$ and will reach negative saturation at time t_2 when area (6+7+8)= area (2+5).

It follows, then, since the voltage applied to reactor A carrying it to positive saturation in the first half-cycle is of the same magnitude as that carrying reactor B in the second-half cycle back to negative saturation, t_2 will be less than t_1 and area (6+7+8)<area (2+3+4).

Also, during the second, half-cycle reactor A, in proceeding from positive saturation, will absorb area 6 from $t=0$ to $t=t_2$ and area 9 from $t=t_2$ to $t=T$. It will be assumed that the area absorbed by the reactor proceeding from saturation, reactor A, again is less than the area absorbed by the reactor which saturated or area(6+9)<area (6+7+8).

This means that, in the third half-cycle, the reactor proceeding toward saturation will begin its excursion nearer saturation than did the corresponding reactor in the second half-cycle. As a result, the firing time t_3 will be less than t_2. Since the times of firing for the second and third half-cycles were different, it is apparent that the transient is not over.

On the other hand, assume that, in the second half-cycle, area (6+9) was equal to area (6+7+8). The reactor proceeding from saturation would have absorbed the same area as the reactor which saturated. It would then begin its excursion toward saturation in the third half-cycle from the same level as did the reactor which saturated in the second half-cycle. As a result, t_3 would be equal to t_2 and the transient would be over.

In summarizing, it may be said, then,

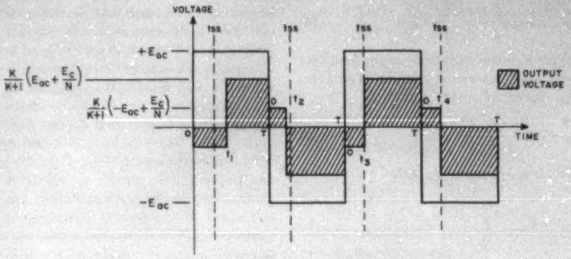


Fig. 4. Output voltage during a transient in mode 2

that in the early stages of the transient, the reactor proceeding from saturation, in a given half-cycle, will absorb fewer volt-seconds than the reactor proceeding toward saturation. As the transient progresses, this difference becomes progressively less and becomes zero when steady state is reached.

It is interesting and important to note that, once the transient has been initiated, the absolute magnitude of the reactor voltages are always identical during similar portions of time. That is, for example, in any half-cycle, a reactor proceeding toward saturation in the nonconducting period has applied to it the voltage $|1/2(E_{ac}+E_c/N)|$. It therefore follows that the average load and control currents are transient only in respect to firing time.

Mode 2: $E_c/N>E_{ac}$

To analyze the circuit for the condition $E_c/N>E_{ac}$, it is desirable to rewrite the period equations for load current, control current, and voltage across the unsaturated reactor, taking into consideration the significance of the condition $E_c/N>E_{ac}$. The circuit under consideration is again the circuit of Fig. 1. Once again initial conditions are described such that the alternating voltage has been applied and the reactors are cycling from knee to knee. At time $t=0$, each reactor is at negative saturation. E_{ac} is zero and increasing in a positive direction; see Fig. 4. Period equations will now be developed for the first four periods after the initiation of a transient $t=0$.

In mode 2 it is necessary to understand that, if a transient is initiated at time $t=0$ when reactor A and B are at negative saturation (points 0, Fig. 2), reactor B will be immediately driven further into saturation. To show this, assume that the transient is initiated an instant after $t=0$ when reactor A and B have deviated an incremental amount from negative saturation. After the instant of application of E_c, i_L, and i_c will remain zero. Equations 1 and 2 may then be solved for e_b.

$$e_b=\frac{1}{2}\left(E_{ac}-\frac{E_c}{N}\right)$$

Since, in mode 2, $E_c/N>E_{ac}$, reactor B will be driven back toward negative saturation. Thus, if the transient were started when both reactors were at negative saturation, the voltage which would tend to appear across reactor B would be in a direction to drive it further into negative saturation.

Hence, if a transient is initiated at time $t=0$ with the reactors at points 0, Fig. 2, reactor B is immediately driven further into negative saturation and the operation begins in a conducting period. Reactor B is effectively short-circuited as shown in Fig. 5(A). With the aid of this figure, the circuit equations for the first period may be written

$$E_{ac}=e_a-i_LR_L \tag{21}$$

$$E_c=Ne_a+i_cR_c \tag{22}$$

Since i_L must equal Ni_c for an unsaturated reactor, i_L may be substituted for Ni_c in equations 21 and 22. Then, solving these equations simultaneously yields

$$i_L=\frac{-E_{ac}+E_c/N}{R_L+R_c/N^2} \tag{23}$$

$$i_c=\frac{-NE_{ac}+E_c}{N^2R_L+R_c} \tag{24}$$

$$e_a=\frac{E_{ac}+KE_c/N}{1+K} \tag{25}$$

Fig. 5(A) shows the assumed direction of i_L to be opposite to that of E_{ac}. Since equation 23 shows i_L to be positive, its direction was assumed correctly. Since the voltage e_a applied to reactor A is greater than E_{ac} and since the reactor design voltage is $1/2E_{ac}$, reactor A will saturate at some time t_1 which is less than $T/2$.

The time t_1 at which reactor A reaches positive saturation is now found by setting the volt-time integral corresponding to e_a equal to 1/2 the volt-time integral of the supply voltage E_{ac} from $t=0$ to $t=T$, or

$$\int_0^{t_1}e_adt=\frac{1+AK}{1+K}\int_0^{t_1}E_{ac}dt=\int_0^T\frac{1}{2}E_{ac}dt \tag{26}$$

and solving for t_1 yields

$$t_1=\frac{1+K}{1+AK}\ \frac{T}{2} \tag{27}$$

The instant of the firing of reactor A finds reactor B also saturated.

Instantaneously, load and control currents change toward new values and, since E_{ac} is still positive and both reactors are saturated, i_c changes toward E_c/R_c and i_L becomes positive, rising toward E_{ac}/R_L. The net results of these currents are to drive reactor A farther into positive saturation and to bring reactor B out of negative saturation. After reactor B reaches the knee of its flux current loop, the voltage e_b appears across reactor B. At this time its control ampere-turns must equal its load ampere turns. Fig. 5(B) illustrates the new circuit conditions from time $t=t_1$ to $t=T$. With the aid of this figure the following can be written.

$$E_{ac}=e_b+i_LR_L \tag{28}$$

$$E_c=-Ne_b+i_cR_c \tag{29}$$

Solving for e_b yields

$$e_b=\frac{E_{ac}-KE_c/N}{1+K} \tag{30}$$

and for i_L and i_c yields

$$i_L=\frac{E_{ac}+E_c/N}{R_L+R_c/N^2} \tag{31}$$

$$i_c=\frac{NE_{ac}+E_c}{N^2R_L+R_c} \tag{32}$$

Equations 30, 31, and 32 describe e_b, i_L, and i_c respectively during the last portion of the first half-cycle from t_1 to T. During this interval of time reactor B has deviated from negative saturation by absorbing volt-seconds according to the equation

$$\int_{t_1}^Te_bdt=\frac{1-AK}{1+K}\int_{t_1}^TE_{ac}dt \tag{33}$$

At time $t=T$, the first half-cycle is completed and E_{ac} reverses polarity. At this instant the voltage e_b across reactor B changes direction and is again opposite to E_{ac}. Since E_c/N is assumed to be greater than E_{ac}, e_b is greater than E_{ac} and load current will continue to flow in the same direction. Reactor A will thus remain in positive saturation. Figure 5(C) illustrates these conditions and the following voltage and current equations may be written

$$E_{ac}=e_b-i_LR_L \tag{34}$$

$$E_c=Ne_b+i_cR_c \tag{35}$$

Solving simultaneously equations 34 and 35, making use of the relation $i_L=Ni_c$ yields

$$e_b=\frac{E_{ac}+KE_c/N}{1+K} \tag{36}$$

$$i_L=\frac{-E_{ac}+E_c/N}{R_L+R_c/N^2} \tag{37}$$

$$i_c=\frac{-NE_{ac}+E_c}{N^2R_L+R_c} \tag{38}$$

Voltage e_b from 0 to t_2 is in a direction to change the flux of reactor B back to negative saturation. Equation 33 shows the volt-second integral absorbed by reactor B, during the first half-cycle, in deviating from saturation. The time t_2 when reactor B will again saturate is found by equating this integral with the integral of e_b from $t=0$ to $t=t_2$

$$\int_0^{t_2} e_b dt=\frac{1+AK}{1+K}\int_0^{t_2} E_{ac}dt =\frac{1-AK}{1+K}\int_{t_1}^{T} E_{ac}dt \tag{39}$$

Solving equation 39 for t_2 in terms of t_1 yields

$$t_2=\frac{1-AK}{1+AK}T-\frac{1-AK}{1+AK}t_1$$

At time t_2 reactor B saturates while E_{ac} is still negative. Load current reverses polarity and flows in the direction of E_{ac}. Current i_L now drives reactor A out of saturation and again $i_L=Ni_c$. The circuit equations for Fig. 5(D) are

$$E_{ac}=e_a+i_LR_L \tag{40}$$

$$E_c=-Ne_a+i_cR_c \tag{41}$$

The circuit voltage and currents are

$$e_a=\frac{E_{ac}-KE_c/N}{1+K} \tag{42}$$

$$i_L=\frac{E_{ac}+E_c/N}{R_L+R_c/N^2} \tag{43}$$

$$i_c=\frac{NE_{ac}+E_c}{N^2R_L+R_c} \tag{44}$$

Equation 30 shows the voltage e_b applied to reactor B as it is deviating from saturation from t_1 to T. Equation 42 shows the voltage e_a applied to reactor A as it is deviating from saturation (t_2 to T). These equations show that the reactor deviating from saturation from t_n to T has the voltage applied to it

$$\frac{E_{ac}-KE_c/N}{1+K}=\frac{1-AK}{1+K}E_{ac} \tag{45}$$

Similarly, equations 25 and 36 show the voltage on the reactor proceeding toward saturation, in any half-cycle, from 0 to t_n is

$$\frac{E_{ac}+KE_c/N}{1+K}=\frac{1+AK}{1+K}E_{ac} \tag{46}$$

It may also be said that a reactor deviating from saturation in an $(n-1)$th half-cycle has a voltage across its load winding equal to

$$\frac{1-AK}{1+K}E_{ac}$$

and in the following nth half-cycle has the voltage

$$\frac{1+AK}{1+K}E_{ac}$$

Since the volt-second integrals corresponding to these two voltages must be equal, the general expression may be written

$$\frac{1-AK}{1+K}\int_{t_{n-1}}^{T} E_{ac}dt=\frac{1+AK}{1+K}\int_0^{t_n} E_{ac}dt$$

Solving for t_n in terms of t_{n-1} yields

$$t_n=\frac{1-AK}{1+AK}T-\frac{1-AK}{1+AK}t_{n-1} \tag{47}$$

From this general expression $t_2, t_3 \ldots t_n$ may be written in terms of t_1.

$$t_2=\frac{1-AK}{1+AK}T-\frac{1-AK}{1+AK}t_1$$

$$t_3=\frac{1-AK}{1+AK}T-\left[\frac{1-AK}{1+AK}\right]^2T+\left[\frac{1-AK}{1+AK}\right]^2t_1$$

$$t_4=\frac{1-AK}{1+AK}T-\left[\frac{1-AK}{1+AK}\right]^2T+\left[\frac{1-AK}{1+AK}\right]^3T-\left[\frac{1-AK}{1+AK}\right]^3t_1$$

Letting

$$\frac{1-AK}{1+AK}=B;\ t_n=(-B)^{n-1}t_1-T[(-B)^{n-1}+(-B)^{n-2}+(-B)^{n-3}+\ldots+(-B)^1]$$

$$\text{or } t_n=(-B)^{n-1}t_1-\left[\sum_{p=1}^{p=n-1}(-B)^pT\right] \tag{48}$$

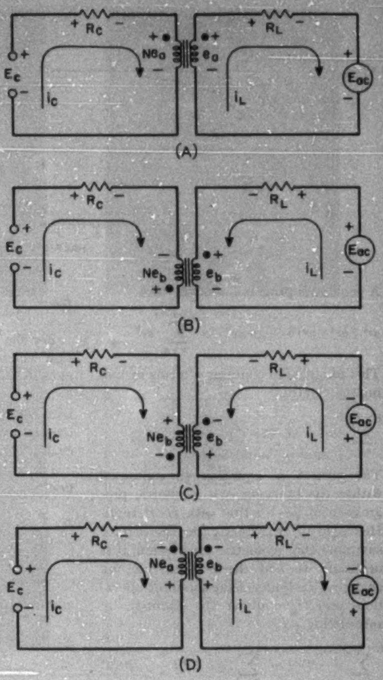


Fig. 5.

A—Circuit conditions from $t=0$ to $t=t_1$
B—Circuit conditions from $t=t_1$ to $t=T$
C—Circuit conditions from $t=T$ to $t=t_2$
D—Circuit conditions from $t=t_2$ to $t=T$

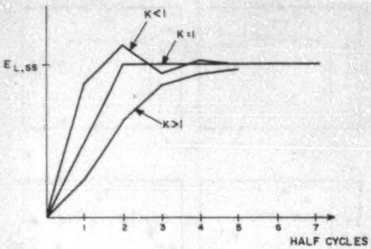


Fig. 6. Output transient responses in mode 1 operation for various values of K

A geometric progression of the form

$$ar^0+ar^1+ar^2+\ldots+ar^{n-1}=\sum_{p=0}^{p=n-1} ar^p$$

This progression contains n terms including the r^0 term.

$$\sum_{p=0}^{p=n-1} ar^p=\frac{a}{1-r}(1-r^n)$$

The portion of equation 48 contained within the brackets is a geometric progression of $n-1$ terms with no r^0 term. Hence, to apply correctly the equation for summing this geometric progression, it is necessary to add the r^0 term, i.e., by changing the limits of the summing operation, and then adjust the quantity by subtracting a T.

$$\sum_{p=1}^{p=n-1}(-B)^pT=\sum_{p=0}^{p=n-1}(-B)^pT-(-B)^0T$$

$$=\frac{T[1-(-B)^n]}{1+B}-T$$

Therefore

$$t_n=\frac{T}{2}[AK-1]\left[\left(\frac{AK-1}{AK+1}\right)^{n-1}-1\right]+\left[\frac{AK-1}{AK+1}\right]^{n-1}t_1$$

From equation 27

$$t_1=\frac{1+K}{1+AK}\,\frac{T}{2}$$

Thus

$$t_n=\left\{(1-AK)+\left[\frac{K(1+A^2K)}{AK+1}\right]\times\left[\frac{AK-1}{AK+1}\right]^{n-1}\right\}\frac{T}{2} \qquad (49)$$

The output voltage in the first half-cycle from $t=0$ to $t=t_1$ is equal to the supply voltage minus the reactor voltage, or

$$E_{ac}-\frac{1+AK}{1+K}E_{ac}$$

Similarly the output voltage in the first half-cycle from $t=t_1$ to $t=T$ can be written as

$$E_{ac}-\frac{1-AK}{1+K}E_{ac}$$

For the purpose of obtaining an equation which describes the progress of the output transient, the output voltage will be assumed to be full wave rectified. The transient equation will then describe the average output voltage over an nth half-cycle.

The average output voltage during the first half-cycle is then

$$E_{L,1,\text{avg}}=\frac{1}{T}\left[-E_{ac}\left(1-\frac{1+AK}{1+K}\right)t_1+E_{ac}\left(1-\frac{1-AK}{1+K}\right)(T-t_1)\right]$$

and in the nth half-cycle

$$E_{L,n,\text{avg}}=\frac{1}{T}\left[-E_{ac}\left(1-\frac{1+AK}{1+K}\right)t_n+E_{ac}\left(1-\frac{1-AK}{1+K}\right)(T-t_n)\right]$$

$$=\frac{E_{ac}}{T}\,\frac{K}{K+1}[-2t_n+(1+A)T] \qquad (50)$$

Substituting t_n from equation 49 into equation 50 yields

$$E_{L,n,\text{avg}}=AKE_{ac}-\frac{K^2E_{ac}(1+A^2K)}{(K+1)(AK+1)}\left[\frac{AK-1}{AK+1}\right]^{n-1} \qquad (51)$$

From equation 47

$$t_n=\frac{AK-1}{AK+1}t_{n-1}-\frac{AK-1}{AK+1}T$$

Therefore

$$t_{ss}=\frac{AK-1}{AK+1}t_{ss}-\frac{AK-1}{AK+1}T$$

$$t_{ss}=(1-AK)\frac{T}{2} \qquad (52)$$

By substituting t_{ss} for t_n in equation 50 the steady-state load voltage $E_{L,ss}$ is obtained

$$E_{L,ss}=AKE_{ac} \qquad (53)$$

The ratio of the average load voltage in the nth half-cycle to $E_{L,ss}$ is found by dividing equation 51 by equation 53

$$\frac{E_{L,n,\text{avg}}}{E_{L,ss}}=1-\frac{K(A^2K+1)}{A(K+1)(AK+1)}\left[\frac{AK-1}{AK+1}\right]^{n-1}$$

or, since

$$E_{L,ss}=AKE_{ac}=K\frac{E_c}{N}$$

$$E_{L,n,\text{avg}}=K\frac{E_c}{N}\left\{1-\frac{K(A^2K+1)}{A(K+1)(AK+1)}\times\left[\frac{AK-1}{AK+1}\right]^{n-1}\right\} \qquad (54)$$

Fig. 4 shows a typical transient for the conditions where

$K<1$

$A>1$

This figure indicates that the transient is nearly completed in three half-cycles.

Conclusions

It was previously stated that it is not obvious from the equations for the control current during the conducting period in mode 1 that the steady-state control current is E_c/R_c. It is now possible, with the relations developed in the transient analysis, to show that for any value of A, i.e., E_c/NE_{ac}, the steady-state control current is E_c/R_c. This, of course, must be true since direct voltage cannot be transformed through the reactors in steady state. If the control current for the conducting period in mode 1 is integrated from the steady-state firing time to the end of the half-cycle T and then divided by the time for the half-cycle T, the resultant current is the average control current for the half-cycle in steady state

$$I_{c,\text{avg}}=\frac{1}{T}\int_{t_{ss}}^{T}\frac{E_c+NE_{ac}}{R_c+N^2R_L}\,dt$$

$$=\frac{1}{T}\int_{\frac{1-AK}{1+A}T}^{T}\frac{E_c}{AR_c}\,\frac{1+A}{1+K}\,dt=\frac{E_c}{R_c}$$

For mode 2 the following relationship can be established

$$I_{c,\text{avg}}=\frac{1}{T}\int_0^{(1-AK)\frac{T}{2}}\frac{-NE_{ac}+E_c}{N^2R_L+R_c}\,dt+\frac{1}{T}\int_{(1-AK)\frac{T}{2}}^{T}\frac{NE_{ac}+E_c}{N^2R_L+R_c}\,dt=\frac{E_c}{R_c}$$

Hence it is seen that the voltage induced into the control circuit from the load circuit merely shapes the control current without affecting its average value. The average value of voltage in the control circuit remains E_c.

As shown in the analysis, two different transients are possible, dependent upon the relative magnitudes of E_c/N and E_{ac}. The resultant transient equations developed from these considerations are certainly different and indicate that the circuit can be capable of two modes of operation. In the analysis the mode described by $E_c/N<E_{ac}$ was called mode 1 and the mode described by $E_c/N>E_{ac}$ was called mode 2. In mode 1 the only circuit parameter which affects the transient is K, while in mode 2 the transient is a function of A and K. It is necessary to understand the significance of the two parameters A and K in order to completely understand the mechanics of operation.

The fundamental factor which determines the mode in which the circuit is operating is the factor A. If E_c/N is less than E_{ac}, the circuit is operating in mode 1; if E_c/N is greater than E_{ac}, the circuit is operating in mode 2. The question now arises as to when the circuit is capable of operating in mode 1 or mode 2. To determine this, a criterion for full output must be established.

In both modes of operation the average load ampere-turns were shown to be equal to the average control ampere-turns. In addition, the steady-state control current was shown to be E_c/R_c. Hence the average output voltage in steady state for any control voltage resulting in full output or less may be written as

$$E_{L,ss}=N\frac{E_c}{R_c}R_L=K\frac{E_c}{N}$$

When full output is reached, E_{ac} will appear across the load resistance R_L for each entire half-cycle. Thus

$$K\frac{E_c}{N}=E_{ac}$$

or $AK=1$.

From this relation it may now be seen that if $K>1$, A may not be greater than 1 before full output is reached. Thus when $K>1$, E_c/N can never be greater than E_{ac} before full output is reached and the circuit is capable of only mode 1 operation. If $K<1$, A may be increased to greater than one before full output is reached. Thus when $K<1$, the circuit may operate in either mode 1 or mode 2. It operates in mode 1 when $A<1$ and mode 2 when $A>1$.

The transient equations are of themselves quite interesting and deserving of additional considerations. For mode 1 the only circuit parameter which affects the transient is K. When K is greater than one, the circuit, after the first half-cycle, approaches steady state in an exponential manner. For $K=1$, the circuit yields steady-state output in one cycle. For a K of less than 1, keeping in mind that A must be less than 1 in order to stay in mode 1, the circuit exhibits a sort of underdamped response. It should be noted, however, that the first half-cycle produces less than the steady-state value of output. The curves of Fig. 6 illustrates the response of the circuit for various values of K in mode 1. Note that the points obtained for this curve represent average output for the corresponding half-cycle.

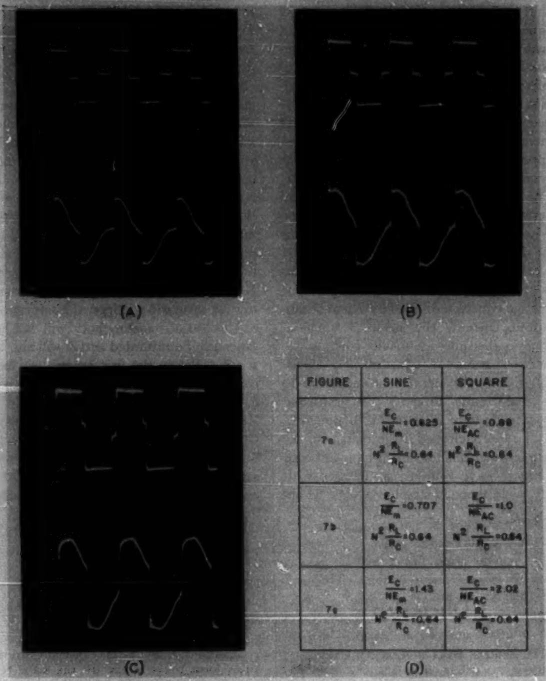


FIGURE	SINE	SQUARE
7a	$\frac{E_C}{NE_m}=0.625$ $N^2\frac{R_L}{R_C}=0.64$	$\frac{E_C}{NE_{AC}}=0.88$ $N^2\frac{R_L}{R_C}=0.64$
7b	$\frac{E_C}{NE_m}=0.707$ $N^2\frac{R_L}{R_C}=0.64$	$\frac{E_C}{NE_{AC}}=1.0$ $N^2\frac{R_L}{R_C}=0.64$
7c	$\frac{E_C}{NE_m}=1.43$ $N^2\frac{R_L}{R_C}=0.64$	$\frac{E_C}{NE_{AC}}=2.02$ $N^2\frac{R_L}{R_C}=0.64$

Fig. 7. Steady-state output current wave forms for typical series-connected saturable-reactor circuit for both square-wave and sine wave supply voltages

A—Output current, mode 1
B—Output current, circuit in transition from mode 1 to mode 2
C—Output current, mode 2
D—Table of circuit conditions for A, B, and C

For mode 2 the transient equation is a function of A and K. However, since AK must be less than one in order to be at less than full output, the term

$$\left(\frac{AK-1}{AK+1}\right)^{n-1}$$

for any value of output less than full output, changes sign on alternate half-cycles. Hence, for mode 2 an underdamped type of response is the only type of response possible, and once again the first half-cycle produces less than steady-state output. Fig. 4 illustrates the response of a circuit in mode 2.

Several interesting conclusions can be drawn by considering a circuit, whose parameters are such that K is less than one, operating in either mode. In mode 1 the steady-state firing time and magnitude of instantaneous output current changes as a function of E_c/N. However, a nonconducting period exists during each half-cycle and the control current is zero for the portion of the half-cycle from 0 to t_{ss}. In mode 2 it is seen from the analysis that the firing time is actually a switching time during which the output current switches between two values. No nonconducting period exists and the control current never goes to zero. The transition from mode to mode occurs when E_c/N becomes equal to E_{ac}. As E_c/N is further increased, the nonconducting period disappears and the load and control currents begin to flow over the entire half-cycle. The firing time becomes the dividing line between two distinct values of currents. When considering the transient equations for mode 1 and mode 2, it is seen that if A is made equal to one in the mode 2 transient equation 54 and the resulting output voltage equation divided by R_L to obtain output current, the mode 2 transient equation becomes identically equal to the mode 1 transient equation 20. Since equation 54 degenerates into equation 20 as A decreases to unity, the transition from mode to mode is seen to be smooth.

The approach used in the analysis of the simple series reactor circuit supplied with a square-wave voltage may be used with any wave shape of supply voltage. However, when sinusoidal excitation is used, for example, the borderline conditions between mode 1 and mode 2 are not clearly defined. For instance, the ratio E_c/Ne_{ac} changes over a half-cycle and may be greater than 1 when the sinusoid is near zero and less than 1 when e_{ac} approaches its maximum value. When the supply voltage is sinusoidal, the circuit will not operate completely in mode 2 until $E_c/N > E_m$, where $e_{ac}=E_m \sin \omega t$. Also, conditions for full output in the proportional range are not as easily described. In mode-2 operation of the circuit, conditions periodically exist when one reactor is saturated and the other is proceeding toward saturation. Upon the saturation of the second reactor, both will instantaneously be saturated and i_c will rise to E_c/R_c. Before the first reactor can become unsaturated, i_L must equal Ni_c, or $e_{ac}/R_L = NE_c/R_c$. In terms of A and K, AK must be equal to or less than unity. Otherwise, load resistance prevents sufficient load current from flowing to unsaturate the first reactor. Thus, when the control voltage is increased, decreasing the firing angle, a point is reached where, at the instant of firing, $AK=1$. Any further increase of control causes $AK>1$ at the instant of firing. The load current will thus be limited by load resistance until e_{ac} increases to make $AK=1$. When AK becomes unity, load current will again be a function of control current only. Thus proportional control will result for low values of control signal up to this boundary condition after which a control characteristic will round into full output. This condition has previously been described as resistance limiting by Storm.[2]

Fig. 7 illustrates steady-state output current wave forms for a typical series-connected saturable-reactor circuit for both square-wave and sine wave supply voltages.

References

1. On the Mechanics of Magnetic Amplifier Operation, Robert A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

2. Magnetic Amplifiers (book), H. F. Storm. John Wiley & Sons, Inc., New York, N. Y., 1955, pp. 122–23.

◆

Discussion

H. C. Bourne, Jr. (University of California, Berkeley, Calif.): The authors have developed an interesting analysis of the series-connected saturable-reactor circuit. The use of a square-wave exciting voltage and a constant d-c control voltage serves the same purpose as the combination of a sinusoidal exciting voltage and a rectified sinusoidal control voltage. If the various sequence of states of operation remain the same, the device is responsive to half-cycle average voltages, as pointed out by the authors. Hence, it is not surprising that the results of mode-1 operation agree with those of Ramey.[1] It is even more interesting to compare the results of mode-1 operation with those obtained by Johannessen.[2] The difference equation 13 developed in this reference with appropriate changes in the symbols reads

$$E_L(n+1)=\frac{1}{N}[E_c(n+1)+E_c(n)]\times \frac{K}{1+K}+E_L(n)\frac{K-1}{K+1} \quad (55)$$

This equation describes the output in the $(n+1)$th half-cycle in terms of the control in the $(n+1)$th and the nth half-cycles and the output in the nth half-cycle. If the special conditions corresponding to equation 20 ($E_c(n)=E_c$, $n>0$ and $E_L(n)=E_c(n)=0$, $n<0$) are substituted, then

$$E_L(n+1)=\frac{1}{N}\frac{2K}{1+K}E_c+E_L(n)\frac{K-1}{K+1} n>0 \quad (56)$$

$$E_L(1)=\frac{1}{N}\frac{K}{1+K}E_c \quad (57)$$

Equations 56 and 57 may be expressed in series form to yield exactly the same results as equation 20. It appears that equation 55 contains all of the information of equation 20 and continues to give a clear picture of the dynamic operation of the circuit. Further, equation 1 is in a form amenable to immediate interpretation into a transfer function, so important from the systems viewpoint.

$$\frac{E_L(z)}{E_c(z)}=G(z)=\frac{\frac{K}{N}(z+1)}{z(K+1)-(K-1)} \text{ in which } z=e^{Ts} \quad (58)$$

With an extension of this technique the discusser has derived the difference equation describing operation in mode 2 as follows

$$E_L(n+1)=\frac{K}{N}[E_c(n+1)+E_c(n)]+\frac{K^2}{N^2E_{ac}(K+1)}[E_c^2(n+1)-E_c^2(n)]+E_L(n)\left[\frac{K}{N}\frac{E_c(n)}{E_{ac}}-1\right] \quad (59)$$

Equation 59 reduces to

$$E_L(n+1)=\frac{1}{N}\frac{2K}{AK+1}E_c+E_L(n)\frac{AK-1}{AK+1} \quad (60)$$

for a step function input and zero initial conditions. With care in determining $E_L(1)$, equation 6 gives the same result as equation 54. Because of the nonlinear character of equation 59 a simple transfer function cannot be obtained and the author's result is probably just as appropriate. However, the applicability of this technique of analysis is again illustrated.

REFERENCES

1. See reference 1 of the paper.

2. Analysis of Magnetic Amplifiers by the Use of Difference Equations, P. R. Johannessen. *AIEE Transactions*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 700–11.

J. F. Ringelman and **A. L. Fenaroli**: The authors wish to thank Dr. Bourne for his interesting discussion. Since the opportunity presents itself, we should also like to commend Mr. Johannessen on his excellent paper (reference 2 of Dr. Bourne's discussion) which introduces a more advanced technique for analyzing magnetic amplifiers.

It is certainly true that the forcing functions chosen by the authors and the forcing functions chosen by Ramey, from the analysis viewpoint, are not radically different. However, it is the authors' contention that the choice of a square-wave supply voltage in conjunction with a d-c signal voltage, aside from simplifying the mathematics involved, aids in visualizing the train of events that occur during a transient. It is also agreed that the defining difference equation as derived by Johannessen powerfully and simply describes the mechanics of operation of the device in a perfectly general sense. The primary purpose of the authors' paper is to describe the mechanics of operation of the circuit with mathematics familiar to the average engineer and, in doing so, derive realistic results and conclusions. The significance of Johannessen's difference equation is not easily appreciated by the novice in the field and it is felt that an appreciation of the period-by-period behavior of the device can best be obtained by investigating the basic nature of the transient. Once this has been firmly established the significance of Johannessen's defining difference equation becomes apparent.

Dr. Bourne is to be commended for writing the transfer function of the magnetic amplifier in a form which places magnetic amplifiers in the correct sphere of analysis. His use of Z in equation 58 suggests a sampled data system of which magnetic amplifiers are a good example. This equation, which is Johannessen's equation 15 rearranged, describes the behavior of the device in mode 1 operation.

Styroflex Aluminum-Sheathed Air-Dielectric Cable

E. J. MERRELL
MEMBER AIEE

A. L. McKEAN
MEMBER AIEE

J. ARBUTHNOTT, JR.
ASSOCIATE MEMBER AIEE

STYROFLEX coaxial cable was developed to meet the demand for a uniform, low-loss cable which could be supplied in long lengths, without joints. Prior to the advent of styroflex, low attenuation in conjunction with high velocity of propagation were attainable only with rigid line or wave guide in short sections joined together.

Developed in its present form by Felten and Guilleaume in Germany,[1] styroflex cable was first introduced in the United States in 1951. Domestic manufacture was begun in 1952 at the Phelps Dodge Habirshaw Division. Extensive use in special applications has proved its versatility throughout the complete frequency spectrum up to 10,000 mc (megacycles).

Styroflex is now being employed in the United States in amplitude and frequency modulation and television broadcasts, Fig. 1; in very-high- and ultrahigh-frequency communications; and in microwave links. The Naval Research Laboratory, University of California, and Massachusetts Institute of Technology have sponsored important installations.

Outstanding properties and advantages of styroflex coaxial cable include the following:

1. Long continuous lengths on reels.
2. Excellent uniformity.
3. Low attenuation.
4. Permanent electrical properties.
5. No radiation.
6. Light weight, high strength.

Typical construction of styroflex cables is shown in Fig. 2. The inner conductor consists of either a solid copper wire or copper tube, which is coaxially supported in a homogeneous aluminum tube by the styroflex tape helix.

The extraordinary properties of styroflex tape make possible the manufacture of a uniform coaxial cable capable of permanently retaining its superior characteristics. Because of the continuous axial support, perfect centering is retained during bending and load cycling, thus assuring retention of initial properties throughout service life. The high-strength helix continuously anchors the inner conductor longitudinally with relation to the outer conductor. Pertinent electrical and physical characteristics of the styroflex line are outlined in Table I.

Manufacture

Styroflex cable is manufactured in long continuous lengths without joints and shipped on reels. A polyethylene jacket is supplied for corrosion protection when the cable is to be installed in underground ducts, directly buried, or for submarine applications.[2]

There are two basic operations in styroflex manufacture. The first operation involves the application of the tape helix on the conductor. A large number of superposed styroflex tapes form the helix. While these tapes are both narrow and thin, 0.100 by 0.005 inch for example, their high modulus of elasticity affords complete stability of the helix and permits a high order of precision in manufacture. Both diameter and pitch are continuously checked, and held to extremely close tolerances as the helical core passes through the taping machine. Styroflex cover tapes enclose the helix, forming a protective binder which enhances dielectric strength.

In the second basic operation, the aluminum sheath is applied by the drawdown or sinking method. The required length of tube, 1,000 feet maximum with present facilities, is drawn out straight from a storage reel and pressure-tested with dry gas. The helical cable core is then drawn in, and finally the tube is drawn through a lubricated sinking die to form a close-fitting sheath over the cable core. Dies are selected to give the exact tube inside diameter specified for characteristic impedance. A high-purity aluminum, 99.7 per cent, is employed as sheathing.

On larger sizes such as $3^{1}/_{8}$ inches, a temporary binding is applied over each press-stop, where local variation in stiffness of the sheath may exist. This is done in order to guard against wrinkling which might occur on subsequent reeling operations. The binding is removed after installations, except when a polyethylene jacket has been applied over it.

Every reel of styroflex must pass rigid physical and electrical factory tests. Attenuation, dielectric strength, characteristic impedance, velocity of propagation, and the voltage standing wave ratio (VSWR) are checked, as are conductivity of copper and aluminum. Each cable length is evacuated, purged, and pressure-tested with dry gas before it undergoes routine tests, and then is pressurized again before the ends are sealed for shipment.

Physical Characteristics

The styroflex cable utilizes pure copper, polystyrene, and aluminum. The tape is a unique flexible form of polystyrene, characterized by stable physical properties. No plasticizer is employed to achieve flexibility, this property being independent of temperature and obtained by special manufacturing techniques. The lack of plasticizer assures constancy of the high tensile and compressive strengths of the styroflex film; the former is over 10,000 psi (pounds per square inch), but most significant is the fact that its elongation at 10,000 psi is less than

---

Paper 56-803, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted October 18, 1955; made available for printing April 30, 1956.

E. J. Merrell, A. L. McKean, and J. Arbuthnott, Jr., are with the Phelps Dodge Copper Products Corporation, Yonkers, N. Y.

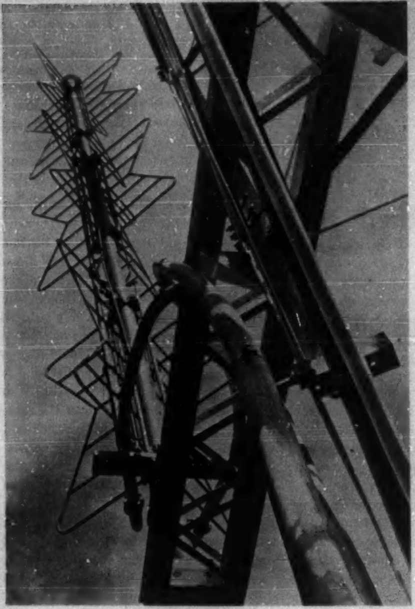


Fig. 1. Styroflex cable serving a television broadcast antenna

3 per cent. This assures complete stability of the helix. The inherent physical strength of styroflex has been demonstrated in one instance by placing a wire between cable core and aluminum sheath during the draw-down operation. Note the impressions in the wire shown in Fig. 3.

The aluminum sheath provides a naturally rugged protective covering which affords high strength and adequate flexibility. While it is recommended that field bending be guided by diameter of shipping reel drum, installed bends in the order of 10 to 15 times the cable diameter are recommended for the smaller cable sizes. However, this is not the practical limit as attested by the usual commercial practice of employing factory "formed-bend" diameters of less than half the recommended figure in the fabrication of broadcast antennas.

Styroflex cable exhibits high tensile strength. The breaking strengths of 7/8- and $3^1/_8$-inch cable were found to be 3,300 and 16,000 pounds, respectively. This is equivalent to a breaking length in the order of 6,000 feet of cable.

Many television installations which employ styroflex are supported solely by the suspension from the top of the transmitter tower, usually by means of a Kellems grip.

Hydrostatic tests to determine collapsing pressure show that standard cables can withstand pressures in the order of 1,200 psi. Hence, without reinforcement, styroflex may be employed as submarine cable to depths of several hundred feet. Service at greater depths is possible with thicker aluminum sheaths, and at very great depths, the deep-sea armored construction, Fig. 4, may be employed.

Electrical Characteristics

Styroflex cable exhibits exceptional electrical uniformity and stability. Being essentially an air-dielectric cable, it inherently exhibits lower attenuation and higher propagation velocity then do equivalent "solid" coaxials.

Its electrical characteristics may be calculated accurately from the properties of its component materials. The general high-frequency formulas (see the appendix) reliably predict cable performance up to approximately 90 per cent of TE_{10} cutoff. In the region of cutoff, a more rigorous analysis, recognizing the helical distribution of the dielectric must be followed, as shown by Griemsmann[3] in an analysis of 1/2-inch styroflex cable.

Calculated attenuation curves for the various standard sizes are given in Fig. 5.

Comprehensive studies of attenuation up to 10,000 mc have been made on 1/2-inch styroflex. Pertinent information for this cable size is tabulated in Table II. The close agreement between calculated and measured attenuation is shown in Fig. 6. The solid line indicates the calculated attenuation for 1/2-inch 50-ohm cable; the points show the record of laboratory measurements.

Initial electrical characteristics are permanent. The effect of temperature cycling upon attenuation is small, and may be estimated from the change in metal resistivities; this amounts to 1 per cent for a 10-degree-Fahrenheit temperature change.

Dielectric loss is a minor factor in the total cable attenuation. This loss is minimized by the low power factor 0.0002, low dielectric constant 2.55, and low "per-cent fill" attributable to the styroflex tape helix.

Between 10 and 100 mc, measurements of input impedance on open-circuited cable of critical length have been used to determine attenuation. Between 100 and

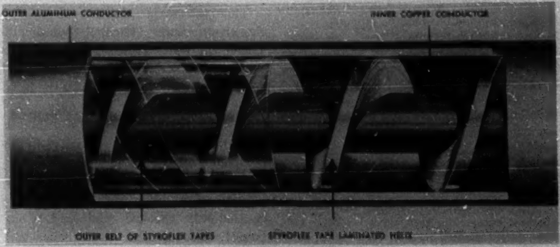


Fig. 2. Cutaway section of styroflex coaxial cable

Vol. 7
Pt. 1
1956

1,000 mc, it has been found satisfactory to employ the substitution method, which entails comparison of the cable with a calibrated attenuator. Above 1,000 mc, spurious reflections from connectors can introduce a serious source of error. In this frequency range, a slotted line in conjunction with sliding short circuit provides a most accurate method; analysis of the data plotted in polar form permits separation of the input reflection from that of the sliding short circuit. The magnitude of the short-circuit reflection is a function of the cable attenuation.

Cable uniformity has been demonstrated in several ways. Measurements of cable standing wave ratio, employing slotted line, indicate values in the order of 1.04 at 2,000 mc. The method of successive resonant frequencies is found to be sensitive for measurements in the 100-mc range. Typical measurements show VSWR les than 1.02. Variation of input impedance over a 20-mc bandwidth provides a qualitative means[4] of uniformity inspection. Fig. 7 shows two oscillograms contrasting the input impedance of styroflex with that of typical *RG* coaxial cable in the very-high-frequency region.

Power ratings have been determined on the basis of experimental heat runs, employing 60-cycle load current, made on 30-foot lengths of unjacketed cable supported horizontally in still air.

At selected temperatures of the inner conductor, ranging from 50 to 85 C (degrees centigrade), measurements of inner conductor, outer conductor, and ambient temperatures were taken, employing thermocouples; the inner conductor temperature was further checked by a resistance bridge. The majority of the tests were made on 50-ohm cable covering 1/2-7/8-, $1^5/_8$- and $3^1/_8$-inch sizes. Several cables of 70- and 77.5-ohm impedance were included. From these tests pertinent thermal transfer characteristics, applicable in this general impedance range, were established as shown in Figs. 8 and 9.

Power ratings were calculated from basic formulas (see Appendix), employing these measured thermal resistances. Independent measurements by others have shown these ratings to be accurate.

The curves of average power rating shown in Fig. 10 were developed on the basis of a maximum inner conductor temperature of 85 C, an ambient of 40 C, and a VSWR of 1.0, and apply to both 50- and 70-ohm cables. Cables having 77.5-ohm impedance are to be derated approximately 15 per cent.

The power curves apply equally to polyethylene-jacketed cables. Labora-

Fig. 3. Demonstration of relative hardness of styroflex tape helix and wire

Table I. Characteristics of 1/2-Inch 50-Ohm Styroflex Cable

Physical		Electrical	
Inner conductor, copper	0.162 inch outside diameter	Power factor, per cent	
		Styroflex tape	0.02
		cable	0.006
Outer conductor, aluminum	0.421 inch inside diameter 0.500 inch outside diameter	Dielectric constant	1.83
Styroflex helix		Velocity of propagation, per cent	87.0
Tape width	0.098 inch		
thickness	0.0023 inch		
pitch	0.55 inch		
Styroflex binder		Capacitance, micro-microfarads per foot	24.0
Tape width	0.787 inch		
thickness	0.0047 inch		
Weight per 1,000 feet	161 pounds	Maximum voltage (peak)	5.4 volts
Installed bending radius	5.0 inches	Leakage current per 1,000 feet	less than 1 microampere at 5 kv
Tensile strength	1,200 pounds	Attenuation decibels per 100 feet	
		100 mc	0.80
		1,000 mc	2.75

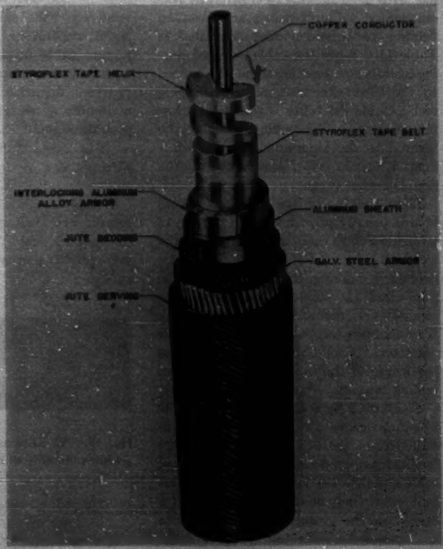


Fig. 4. Typical construction of deep-sea armored styroflex cable

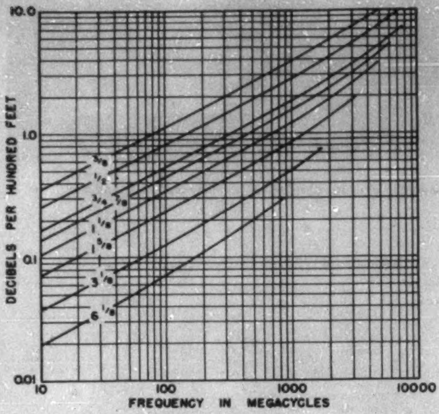


Fig. 5. Attenuation characteristic of styroflex coaxial cables

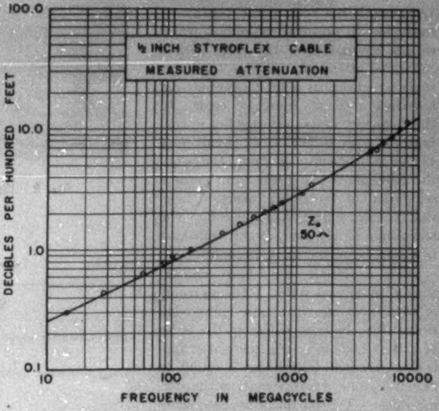


Fig. 6. Measured attenuation of 1/2-inch 50-ohm styroflex cable

tory heat runs actually show about 10-per-cent improvement in power rating due to a change from bright aluminum to black jacket.

Handling and Installation

Installation procedures for styroflex cable follow practices customary in working with lead-sheathed cable. The recommended pulling tensions and installed bending radii for styroflex are listed in Table II. When the cable is pulled during installation, bending should be limited to the drum diameter of the shipping reel. Cable flexibility is retained at low temperatures. The pulling cable may be secured to the aluminum sheath with a long Kellems grip, preferably of the double becket type.

Light braking on the reel flanges with boards usually provides ample tension to straighten cable satisfactorily through the $1^1/_8$-inch size. For larger diameters, necessary braking of the reel may be minimized by employing a simple box straightener, mounted close to the reel. A typical box length is 25 times the cable diameter. A light wipe of mineral oil on the inside of the box will help to reduce pulling tension.

Standard 2 by 8 plank is ample for the $3^1/_8$-inch cable straightener, and construction of the box with carriage bolts facilitates operations. Thus, a length of cable equal to the length of the box plus pulling grip may be straightened against a board on the ground, then swung into the opened box, the top of which is secured by bolting. The straightener may be rigged to traverse the reel as the cable is unwound. Since the box straightener described offers little braking and can be restrained with light tackle, these are relatively simple procedures; see Fig. 11.

Though some styroflex is installed in ducts or buried in the ground, much of it is aerial. The high order of self-support due to the aluminum sheath permits relatively wide spacing of trolleys and aerial suspensions. Steel-reinforced aluminum cable messenger, in conjunction with suitable aluminum lashing wire, provides adequate aerial support and maintains a neutral electrochemical aspect throughout the cable assembly.

Fittings

A complete line of fittings, proved in service, are available for interconnection of styroflex cable, and for connecting styroflex to other types of coaxial cable or equipment. Their flare-type design affords mechanical strength virtually equal to the cable.

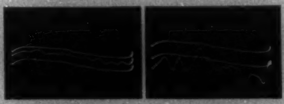


Fig. 7. Oscillograms illustrative of the greater uniformity of styroflex cable compared to RG types

A—1/2-inch 50-ohm styroflex (VSWR, 1.01)
B—Type RG-8/U (VSWR, 1.05)
Frequency band—70–90 mc (A) (B)

Applications

In service, high-frequency cables are required to maintain their electrical properties through a variety of operating conditions. They must be able to resist moisture, oxidation, and chemical attack, and must remain electrically stable under a wide range of temperature cycling. The excellent performance record of styroflex has led, inevitably, to new coaxial applications as well as to improvement in the coaxial components of many present-day systems and equipment.

Styroflex cable has been successfully employed as a radar calibration standard, wherein faithful reproduction of signal, constancy of transit time, and stability of attenuation provide the uniformity demanded of a "standard." Data transmission from remote test sites has been accomplished, efficiently, with this cable. In such capacity, its high propagation velocity and fidelity, plus availability in long reel lengths, greatly facilitates its performance and installation. Styroflex is also being employed as primary communications cable on such projects as the Defense Early Warning (DEW) line.

The continuous styroflex transmission line is particularly adapted for tower installations, both because of simplicity in handling and permanence of electrical characteristics and performance. Although $3^1/_8$-inch cable is commonly used,

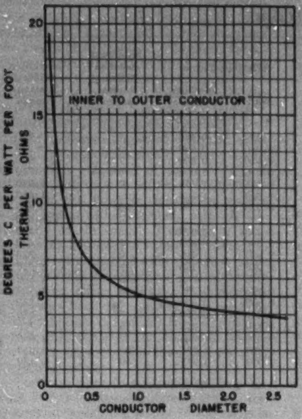


Fig. 8 (left). Thermal resistance of styroflex cable, inner conductor to outer conductor

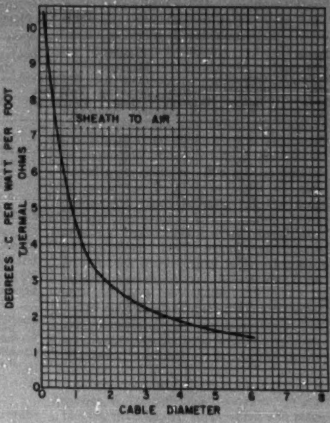


Fig. 9 (right). Thermal resistance of styroflex cable, outer conductor to air

installations of $6^1/_8$-inch styroflex have been made, the latest being the current application at station *WIIC* in Pittsburgh, Pa., where the tower rises 800 feet. This is actually the first $6^1/_8$-inch styroflex transmission line in the United States. Several similar installations are under consideration. Styroflex is also being extensively employed for antenna feed lines in scatter propagation.

Due to its broad-band uniformity, styroflex is admirably suited for multichannel long-line telephone communications. The same characteristic has led to advantageous employment in community antenna television systems where signal transmission, in some instances, embraces a 150-mc bandwidth, starting at about 50 mc. Low attenuation and permanence guarantee an efficient network cable. Absence of radiation, and stability under temperature cycling, enhance its performance. Significant economies are realized through appreciable reduction in the number of system amplifiers, and elimination of the need for cable replacement and maintenance, which justifies the use of styroflex over braided coaxial.

In general, styroflex has proved its efficiency and reliability in virtually every type of coaxial application.

Appendix I. Computation of Electrical Characteristics of Styroflex Coaxial Cable

Nomenclature

d_1 = outer diameter of inner conductor
d_2 = inner diameter of outer conductor
w = width of styroflex tape
p = lay of styroflex helix
Z_0 = characteristic impedance
k_r = dielectric constant of cable
k_s = dielectric constant of styroflex (2.55)
α_t = total attenuation in decibels per 100 feet
α_m = total metal losses in decibels per 100 feet
α_c = inner conductor loss in decibels per 100 feet
α_d = dielectric loss in decibels per 100 feet
$\tan\delta_s$ = loss tangent of styroflex (2×10^{-4})
$\tan\delta_r$ = loss tangent of cable
F_{mc} = frequency in megacycles
P = power rating in watts in still air
Δ_T = temperature rise of the inner conductor above ambient
R_{THs} = thermal resistivity between sheath and air in degrees centigrade per watt-foot
R_{THc} = thermal resistivity between inner conductor and sheath in degrees centigrade per watt-foot

Formulas

Characteristic impedance (ohms):

$$Z_0 = \frac{60}{\sqrt{k_r}} \ln\left(\frac{d_2}{d_1}\right)$$

Relative dielectric constant:

$$k_r = 1 + (k_s - 1)\frac{w}{p}\sqrt{\frac{1+p^2}{\pi 2_{d_1 d_2}}}$$

Total attenuation (decibels per 100 feet);

$$\alpha_t = \alpha_m + \alpha_d$$

Total metal loss (decibels per 100 feet);

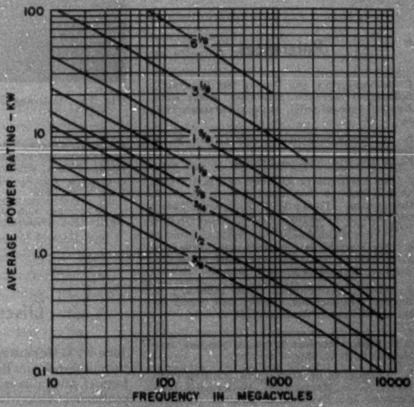


Fig. 10. Power ratings of styroflex cable

Table II. Nominal Characteristics of Styroflex Coaxial Cables

Characteristic Impedance, Ohms	Over-all Diameter, Inches	Dielectric Constant	Capacity, Magnetomotive Forces per Foot	60-Cycle Peak Test Voltage, Kv	Inner Conductor		Outer Conductor		Habiriene Jacket		Minimum Installed Bending Radius, Inches	Maximum Pulling Tension, Pounds
					Wall	Outside Diameter	Inside Diameter	Outside Diameter	Wall	Outside Diameter		
50	3/8	1.40	24	2.2	solid	0.112	0.296	0.375	0.030	0.435	4	210
	1/2	1.34	24	3.4	solid	0.162	0.421	0.500	0.050	0.600	5	270
	3/4	1.24	23	5.0	solid	0.251	0.632	0.750	0.050	0.850	7.5	620
	7/8	1.23	22	6.0	solid	0.300	0.758	0.875	0.070	1.015	10	770
	1 1/8	1.23	22	8.0	0.051	0.400	1.007	1.125	0.070	1.265	17	1,040
	1 5/8	1.20	22	11.0	0.055	0.591	1.472	1.625	0.070	1.765	25	1,850
	3 1/8	1.18	22	19.0	0.071	1.157	2.850	3.125	0.070	3.265	50	6,450
	6 1/8	1.16	22	45.0	0.118	2.25	5.512	6.125	...	...	90	28,000
70	1/2	1.29	17	3.4	solid	0.114	0.421	0.500	0.050	0.600	5	270
	3/4	1.25	16	5.0	solid	0.173	0.632	0.750	0.050	0.850	7.5	620
	7/8	1.24	16	6.0	solid	0.209	0.758	0.875	0.070	1.015	10	770
	1 1/8	1.24	16	8.0	solid	0.276	1.007	1.125	0.070	1.265	17	1,040
	1 5/8	1.25	16	11.0	0.051	0.403	1.472	1.625	0.070	1.765	25	1,850
	3 1/8	1.20	16	19.0	0.071	0.800	2.850	3.125	0.070	3.265	50	6,450
77.5	1/2	1.30	15	3.4	solid	0.099	0.421	0.500	0.050	0.600	5	270
	3/4	1.26	15	5.0	solid	0.149	0.632	0.750	0.050	0.850	7.5	620
	7/8	1.25	15	6.0	solid	0.180	0.758	0.875	0.070	1.015	10	770
	1 1/8	1.25	15	8.0	solid	0.240	1.007	1.125	0.070	1.265	17	1,040
	1 5/8	1.24	15	11.0	solid	0.354	1.472	1.625	0.070	1.765	25	1,850
	3 1/8	1.21	15	19.0	0.071	0.690	2.850	3.125	0.070	3.265	50	6,450

$$\alpha_m=\frac{0.0174}{\ln\frac{d_2}{d_1}}\left[\frac{0.414}{d_1}+\frac{0.525}{d_2}\right]\sqrt{k_r F_{mc}}$$

Total dielectric loss (decibels per 100 feet):

$$\alpha_d=2.78\tan\delta_r\sqrt{k_r}F_{mc}$$

Dielectric loss tangent:

$$\tan\alpha_r=\frac{k_s}{k_r}\,\frac{(k_r-1)}{(k_s-1)}\tan\delta_s$$

Power rating in still air (watts):

$$P=\frac{434\Delta T}{\left[\alpha_i R_{THs}+\left(\alpha_c+\frac{\alpha_d}{2}\right)R_{THc}\right]}$$

Inner conductor loss (decibels per 100 feet);

$$\alpha_s=\frac{0.0174}{\ln\frac{d_2}{d_1}}\left[\frac{0.414}{d_1}\right]\sqrt{k_r F_{mc}}$$

Fig. 11. Installing $3^1/_8$-inch styroflex cable transmission line on television broadcast tower

References

1. Manufacture and Use of the Electrical Insulating Material Styroflex, Heinz Horn. *Kunstoffe* (special pub.), Munich, Germany, 1940, pp. 53–57.

2. Aluminum-Sheathed Power Cable, W. A. Del Mar, E. J. Merrell. *AIEE Transactions*, vol. 74, pt. III, June 1955, pp. 292–304.

3. An Approximate Analysis of Coaxial Line with a Helical Dielectric Support, J. W. E. Griemsmann. *Transactions*, Microwave Theory and Techniques, Institute of Radio Engineers, New York, N. Y., vol. MTT-4, no. 1, 1956.

4. Applications of a Sweep Signal Generator, F. G. Marble. *The Notebook*, Boonton Radio Corporation, Boonton, N.J., vol. 7, 1955.

5. Report on Styroflex, G. M. Dinnick, J. K. Burton, J. R. Ambrose. *Report 4160*, Naval Research Laboratory, Washington, D. C.

6. Microwave Transmission Circuits (book), G. L. Ragan. McGraw-Hill Book Company, Inc., New York, N. Y., 1948.

7. Techniques of Microwave Measurements (book), C. G. Montgomery. McGraw-Hill Book Company, Inc., New York, N. Y., 1947.

8. Measuring the Efficiency of a Superheterodyne Converter by the Input Impedance Circle Diagram, H. A. Wheeler, D. Dettinger. Wheeler Monograph No. 9, Wheeler Laboratories, New York, N Y., 1949.

9. Electrical and Mechanical Characteristics of the Styroflex-Helix Coaxial Cable, K. S. Wyatt, L. Bellingham. Conference on Electrical Insulation, National Research Council, Washington, D. C., 1951.

10. A Method of Making Radar Self-Calibrating, E. A. Wolff. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 42, no. 10, Oct. 1954.

◆

Discussion

John W. E. Griemsmann (Polytechnic Institute of Brooklyn, Brooklyn, N. Y.): The form of this cable is of interest in that the continuous dielectric provides for broadbanding not possible with lines having spaced dielectric supports. This is advantageous when long lengths or high multiplicity of supports are required. The advantage in attenuation over solid dielectric cables is obvious and again appears important when long lengths are involved.

Analytically, any cable is subject to impedance instability resulting from minute local variations in characteristic impedance along the length of the cable. This effect, for all cables, is accentuated by long lengths and low attenuation. While a comparison of impedance stability has been made in the paper for the very-high-frequency region, a comparison of impedance uniformity with that of corresponding solid dielectric cable at higher frequencies would provide useful information. Because of the intensity of gaseous ionization possible at higher frequencies, the 60-cycle breakdown values would appear to indicate only basic electric integrity of the cable. Here again, comparison of breakdown values at radio frequencies with that of solid dielectric cable would provide useful information relating to high-power pulse applications. Important differences between styroflex and solid dielectric flexible cables are that styroflex is semiflexible and that the connector assembly job is more complex than that for solid dielectric cables.

The more rigorous analysis given in reference 3 of the paper stipulates a critical frequency at which the power in the cable is directed parallel to the helix. The values of critical frequency computed from the theory correspond approximately to values of cutoff of the first unsymmetrical mode, assuming a homogeneous dielectric and using the effective value of dielectric constant as indicated in Table III.

Table III. Values of Critical Frequency

Line Size, Inches	Critical Frequency, Kilomegacycles	Cutoff, First Unsymmetrical Mode, Kilomegacycles
3/8	14.02	15.7
1/2	11.31	11.17
7/8	6.85	6.42
$1^{1}/_{4}$	3.82	3.28

While the theory shows a calculable difference from the more elementary theory of the paper at one-half critical frequency, appreciable changes in attenuation and wave length (in cable) does not appear to take place for frequencies lower than about 90 per cent of critical frequency. Measured values of the departure of attenuation and wave length (in cable) from those given by the low-frequency values have been found to be in agreement with theory.

Tests made at microwave frequencies indicate that the recommended values of minimum bending radii will not give rise to significant additional reflections.

E. J. Merrell, A. L. McKean, and **J. Arbuthnott, Jr.:** Dr. Griemsmann's comments and suggestions are sincerely appreciated. We would like to supplement the electrical measurements given in the paper and briefly comment on some points of discussion.

Recent laboratory measurements of uniformity made over 50-mc bandwidths up to 4,000 mc show the VSWR of 1/2-inch 50-ohm styroflex to be about 1.04 at 2 kilomegacycles, 1.06 at 3 kilomegacycles, and 1.08 at 4 kilomegacycles. The measurements indicated that the sources of nonuniformity were random rather than uniformly spaced small discontinuities. Cable test lengths were 1,000 feet.

With regard to breakdown voltage of styroflex, it should be made clear that the 60-cycle test voltages listed in Table I are withstand voltages carrying a factor of safety; they are not breakdown values. Breakdown at 60 cycles may be estimated by employing a limiting voltage gradient of 70 volts per minute. This applies equally at megacycle frequencies based on continuous a-c characteristics. If the peak voltage of a pulse of high-frequency power does not exceed this gradient, breakdown will not occur. Under certain conditions, a greater peak gradient may apply, when considering shorter pulses or those having lower pulse repetition frequencies.

The semiflexible characteristic of styroflex aluminum-sheathed cable is recognized in handling, hence we have indicated that recommended installation procedures follow lead-sheathed cable practice.

Where maximum flexibility is required in service, a "styroflexible" construction employing silver-plated copper braids in place of aluminum has been made available in several smaller sizes. A polyethylene jacket is provided over all.

Styroflex connectors employ a basic flare-type design, and thus afford a permanent, sealed connection. Good electrical design in conjunction with relatively simple assembly has been achieved, so that the operation of installing connectors requires only normal skill, and follows straightforward installation procedures.

The Analytical Design and Evaluation of Electromagnets

M. J. KELLY
NONMEMBER AIEE

J. E. WALLACE
NONMEMBER AIEE

Synopsis: This paper describes a method for the accurate computation of the operating characteristics of elecromagnets, such as used in relays and other electromechanical devices. The dynamic solution involves the use of techniques applicable to digital computers. The evaluation of the computer solutions offers a means of obtaining the optimum magnet design for a given device. The method eliminates the need for building successive models for testing and evaluation.

THE DESIGN of electromagnetic devices, such as relays, has heretofore been based on empirical procedure, with the efficiency of the device largely a measure of the skill of the designer. The finished design of an electromagnet usually was the result of the construction and testing of several models. Even then it is doubtful if maximum efficiency was approached. It is the intent of this paper to describe a method whereby mathematical theory is used to reduce this cut-and-try procedure to a minimum.

Mathematical calculation of the operating characteristics of an electromagnet previously was considered extremely difficult, if not impossible, because of the large number of interdependent variables that are involved. However, by means of digital computer techniques, an analytical solution has been evolved that has given a fairly high degree of accuracy for the cases studied.

The method of solution described here requires that each variable be mathematically described as accurately as possible. By the use of difference calculus each expression may be solved individually and the results applied to the solution of the next expression. The final solution is obtained by reiterative calculation. The results of the mathematical solution will not be exact; however, they will be well within the range of the physical variables encountered in a model due to material, air gap, force, and other variations inherent in a mechanical system.

For purposes of simplification all forces are equated to act at the center of the main working air gap. This includes the inertia forces present during the dynamic portion of the operate cycle, the return spring force, and the mechanical force

---

Paper 56-787, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted June 16, 1955; made available for printing May 23, 1956.

M. J. Kelly and J. E. Wallace are with International Business Machines Corporation, Endicott, N. Y.

The authors wish to express their appreciation to Margaret A. McCormack and Bernard M. Tostanoski of the Endicott IBM Scientific Computation Laboratory for their assistance in the application of the problem to digital computer methods.

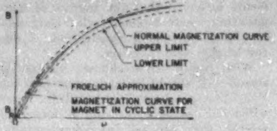


Fig. 1. Froelich approximation of magnetization curve

which may be represented by a contact load or other load.

The mathematical solution is taken in two stages. The first stage is the portion of the operate cycle required for the flux to rise to the point where sufficient magnetic force is generated to overcome the initial resistance of the mechanical system. That is the point in time at which motion is incipient. This calculation is fairly simple and may be done with a desk calculator or a slide rule. The first stage is necessary to provide the constants to be used for the dynamic solution of the second stage. The second stage involves the reiterative solution of the sequence of equations 27 to 38 and 54 to 68.

Discussion

NOMENCLATURE

Some of the variations in subscripts in the following symbols are not given here, but explained where they are used in the text.

Time Symbols

t = time, seconds (sec)
t_0 = time at which motion is incipient, sec
Δt = increment of time used for calculation of the second stage of the operation, sec

Magnetic Symbols

A = pole-face area, square inches (in.^{2})
A' = equivalent pole-face area, in.^{2}
X = length of gap at center of pole face, inches (in.)
l_t = length of iron path in magnet yoke, in.
l_c = length of iron path in magnet core, in.
l_a = length of iron path in armature, in.
l_H = length of assumed air path at hinge, in.
l_r = length of residual air gap, in.
a_t = cross-sectional area of yoke, in.^{2}
a_c = cross-sectional area of core, in.^{2}
a_a = cross-sectional area of armature, in.^{2}
a_H = cross-sectional area of hinge gap, in.^{2}
a_r = cross-sectional area of residual gap, in.^{2}
N = number of coil turns
NI = ampere-turns
$\mathcal{R}$ = reluctance, NI/maxwells
$\mathcal{R}_l$ = leakage reluctance, NI/maxwells
$\mathcal{R}_a$ = reluctance of armature, NI/maxwells
$\mathcal{R}_c$ = reluctance of core, NI/maxwells
$\mathcal{R}_t$ = reluctance of yoke, NI/maxwells
$\mathcal{R}_H$ = reluctance of hinge gap, NI/maxwells
$\mathcal{R}_r$ = reluctance of residual gap, NI/maxwells
$\mathcal{R}_o$ = closed gap reluctance except R_a, NI/maxwells
$\mathcal{R}_g$ = reluctance of magnet exclusive of leakage reluctance, NI/maxwells

$$\mathcal{R}_g=\mathcal{R}_o+\mathcal{R}_a+\frac{0.313\ X}{A'}$$

$\dfrac{0.313\ X}{A'}$ = reluctance of equivalent working gap NI/maxwells
$\mathcal{R}_t$ = total reluctance, NI/maxwells
ϕ = flux, kilomaxwells
ϕ_g = flux passing through working air gap, kilomaxwells
ϕ_l = leakage flux, kilomaxwells
ϕ_t = total flux, kilomaxwells
ϕ_r = residual flux, kilomaxwells
μ = permeability, gauss/oersteds
f = force due to magnetic pull, pounds (lbs)
B = flux density, kilomaxwells/in.^{2}
H = field intensity, NI/in.
C_1 = Froelich equation constant
C_2 = Froelich equation constant
$\mathcal{P}$ = permeance

Electrical Symbols

E = steady-state voltage, volts
I = steady-state current, amperes
e = instantaneous voltage, volts
i = instantaneous current, amperes
R = resistance, ohms
L = inductance, henry
ρ = resistivity, ohms-centimeter
ϵ = energy, joules

Mechanical Symbols

F_A = force available to accelerate armature, lbs
F = mechanical force, lbs
F_0 = equivalent mechanical force acting at the center of the pole face to start motion, lbs
P = equivalent mechanical force acting at the center of the pole face, generated due to armature motion $P=f(S)$, lbs
I' = moment of inertia, in.-lbs-sec^2
r = radius, in.
M = mass, lbs sec^2/in.
M' = equivalent mass acting at the center of the pole face, lbs sec^2/in.
S = displacement at the center of the pole face, in.
V = velocity of the armature at the center of the pole face, in./sec
K = drop-out spring rate, lbs/in.

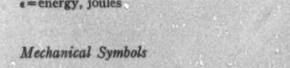


Fig. 2. Basic magnet structures

1—Two leg
2—Three leg
3—E-core

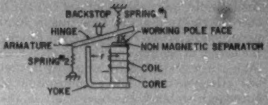


Fig. 3. Diagram of simple 2-leg magnet

Note that throughout the text the subscript o is used to denote the value of a quantity at the time (t_o) when motion is incipient; for example, X_o, ϕ_{go}, etc.

MAGNETIZATION OF MATERIALS

The problem of approximating the magnetization curve by a mathematical expression was solved by Froelich and will be used in this analysis. The Froelich equation is a hyperbolic approximation with the origin at zero. The expression is as follows:

$$B=\frac{C_1H}{C_2+H} \tag{1}$$

Fig. 1 shows the relationships discussed, from which it may be seen that the Froelich approximation is of sufficient accuracy in the usual magnet design.

MAGNET DESIGN

The types of magnet structure most frequently encountered are illustrated in Fig. 2.

Each magnet type offers certain advantages. The type of magnet selected is left to the designer's discretion. In general, the type of application, the load characteristics, and the direction of desired motion will determine the type of magnet structure to be used.

Calculations for the different magnet types differ only in detail. The 2-leg core and the E core can be equated equal and the same solution used for either unit. The 3-leg core must have a modified solution due to the unequal flux distribution in the outer legs.

METHOD OF SOLUTION

So that the nature of the problem may be understood, the simple 2-leg magnet shown in Fig. 3 and the factors which enter into its operation will be considered.

In order to arrive at a solution for the operate characteristics of the sample magnet, the initial conditions must be evaluated. The procedure is as follows:

1. The total inertia of the moving system must be evaluated and converted to the equivalent mass at the center of the pole face by means of the expression:

$$M'=\frac{I'}{r^2} \tag{2}$$

Vol. 75
Pt. 1
1956

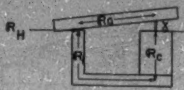


Fig. 4. Reluctances of the 2-leg magnet

$\mathcal{R}_a$ = reluctance of armature
$\mathcal{R}_c$ = reluctance of core
$\mathcal{R}_t$ = reluctance of yoke
$\mathcal{R}_H$ = reluctance of hinge
$\mathcal{R}_r$ = reluctance of residual gap
$\mathcal{R}_l$ = leakage reluctance

$\frac{0.313X}{A}$ = reluctance of working gap

2. The initial spring forces must be translated to act at the center of the pole face by taking moments. This force will be designated by F_o.

3. The force F_o is the mechanical force that must be overcome for motion to take place. Therefore, at the time motion is incipient, the magnetic force f_o must be equal to F_o; that is, at t_o, $f_o=F_o$, where t_o is the time where motion starts and is the finish of the first stage of the calculation.

$$f_o=\frac{0.0139\,\phi_{go}^2}{A} \tag{3}$$

then

$$\phi_{go}=\left(\frac{F_o\,A}{0.0139}\right)^{1/2} \tag{4}$$

This will give the value of flux ϕ_{go} required to begin motion.

4. The total reluctance of the magnetic circuit must be calculated. The reluctances of the 2-leg magnet are indicated in Fig. 4.

The following conditions apply: As reluctance of two mating metal surfaces indicates an equivalent spacing of 0.0015 to 0.002 inch, the length of the gap at the hinge end is arbitrarily taken at 0.002 inch if the armature rests on the core; and the reluctance of the core and yoke are considered to be constant at the value of ϕ_{go} derived for the particular magnet. (Care must be taken to be sure that the yoke and core operate below the knee of the magnetization curve for maximum value of flux.)

The approximate equivalent circuit for the reluctance of the 2-leg magnet is seen in Fig. 5.
Where

$$\mathcal{R}_o=\mathcal{R}_c+\mathcal{R}_t+\mathcal{R}_h+\mathcal{R}_r \text{ (assumed constant)} \tag{5}$$

$$\mathcal{R}_t=\frac{\mathcal{R}_l\left(\mathcal{R}_a+\mathcal{R}_o+\frac{0.313\,X}{A}\right)}{\mathcal{R}_l+\mathcal{R}_a+\mathcal{R}_o+\frac{0.313\,X}{A}} \tag{6}$$

As the reluctance of the iron in the circuit is a function of ϕ, the expression for R is expressed in terms of the Froelich constants and is given as follows:

$$\mathcal{R}=\frac{C_2l}{aC_1-\phi_g}\times10^{-8} \tag{7}$$

The procedure for the calculation of reluctances is as follows:

$$\mathcal{R}_H=\frac{0.313l_H}{a_H} \tag{8}$$

$$\mathcal{R}_r=\frac{0.313l_r}{a_r} \tag{9}$$

$$\mathcal{R}_c=\frac{C_2l_c}{a_cC_1-\phi_{go}}\times10^{-8} \tag{10}$$

$$\mathcal{R}_t=\frac{C_2l_1}{a_1C_1-\phi_{go}}\times10^{-8} \tag{11}$$

$$\mathcal{R}_o=\mathcal{R}_H+\mathcal{R}_r+\mathcal{R}_c+\mathcal{R}_t \tag{12}$$

$$\mathcal{R}_a=\frac{C_2l_a}{a_aC_1-\phi_{go}}\times10^{-8} \tag{13}$$

$$\mathcal{R}_g=\mathcal{R}_o+\mathcal{R}_a+\frac{0.313\,X}{A} \tag{14}$$

The leakage reluctance $\mathcal{R}_l$ is now solved.[1]
The total open gap reluctance of the magnet is

$$\mathcal{R}_t=\frac{\mathcal{R}_l\mathcal{R}_g}{\mathcal{R}_l+\mathcal{R}_g} \tag{15}$$

5. The initial inductance of the magnet may now be calculated. It will be seen from the foregoing calculations that the greater part of the total magnet reluctance resides in the working and residual air gaps. For this reason, the indutance of the magnet may be considered to be constant until the armature starts to move and the air gaps diminish.

As indicated previously the presence of residual flux in the iron and hence in the air gap, causes a slight error in the initial value of L. The following expression is used for the calculation of L_o.

$$L_o=\frac{N^2\times10^{-8}}{\mathcal{R}_t} \tag{16}$$

6. The value of current which is required to produce ϕ_{go} is:

$$i_o=\frac{\mathcal{R}_{go}\phi_{go}}{N}\times10^8 \tag{17}$$

7. The time t_o at which motion is incipient is:

$$t_o=\frac{L_o}{R}\ln\left(\frac{E}{E-Ri_o}\right) \tag{18}$$

8. This ends the calculation for the first stage of the relay operate cycle.

9. The second stage of the operate solution is accomplished by a digital computer. In addition to the previous calculations the following parameters must be defined.

The mechanical force which is developed by the motion of the armature is designated by $P=f(S)$. $f(S)$ may be any complicated force function which may be approximated by a series of spring constants multiplied by the displacement. The force P is translated by means of moments to the center of the main working air gap. For example: In the magnet shown in Fig. 3 assume spring no. *1* has a nonlinear rate and spring no. *2* has a linear rate. The resultant force at the radius r is shown by the force-displacement diagram in Fig. 6.

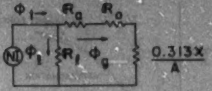


Fig. 5. Equivalent circuit of the 2-leg magnet

$P=f(S)$ may be described as follows:

$P=K_1S \quad 0<S<S_1$

$P=K_2S \quad S_1<S<S_2$

$P=K_3S \quad S_2<S<S_3$

Kn = equivalent spring rate when n = 1,2,3, etc.

If there is a sudden or gradual change in the equivalent mass of the moving system during the armature motion, the mass may be described in a similar manner.

10. It has been previously stated that the solution for the *E*-core may be obtained by the same methods as for the 2-leg core. To do this the equations for the *E*-core must be made equivalent to those of a hypothetical 2-leg core.

Consider the magnetic circuit illustrated in Fig. 7.

The leakage reluctance ($\mathcal{R}_l$) is now solved.[1]

The following relationships will hold if $a_2=a_3$

$$\mathcal{R}_t=\frac{C_2l_t}{2a_2C_1-\phi_{go}}\times10^{-8} \tag{19}$$

$$\mathcal{R}_c=\frac{C_2l_c}{a_cC_1-\phi_{go}}\times10^{-8} \tag{20}$$

$$\mathcal{R}_a=\frac{C_2l_a}{2a_aC_1-\phi_{go}}\times10^{-8} \text{ if } a_{a1}=a_{a2} \tag{21}$$

$$\mathcal{R}_r=\left(\frac{0.313\,l_{r2}}{A_2+A_3}\right)+\left(\frac{0.313\,l_{r1}}{A_1}\right) \tag{22}$$

In this expression l_{r2} is usually taken as 0.0015 to 0.002 inch depending on the degree of flatness of the mating surfaces.

l_{r1} is the length of the residual gap at the center of the working pole face.

$$\mathcal{R}_o=\mathcal{R}_t+\mathcal{R}_c+\mathcal{R}_r$$

The force developed in the magnet shown in Fig. 7 is generated in the main working gap and at the outer legs of the magnet. As the force may be directly related to the total change in reluctance in the magnet and X_1 and X_2 have a proportional relationship, an expression may be derived which relates the working gap reluctances to X_1 as illustrated in Fig. 8.

Assume $X_2=X_3$ and $A_2=A_3$

$$\mathcal{R}_{x2}=\frac{0.313X_2}{2A_2}$$

$$\mathcal{R}_{x1}=\frac{0.313X_1}{A_1}$$

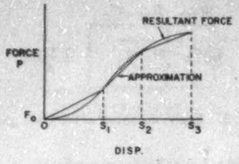


Fig. 6. Static force displacement diagram

$$\frac{X_1}{r_1}=\frac{X_2}{r_2}$$

$$X_2=\frac{r_2X_1}{r_1}$$

$$\mathcal{R}_x=\mathcal{R}_{x1}+\mathcal{R}_{x2}$$

$$=\frac{0.313X_2}{2A_2}+\frac{0.313X_1}{A_1}$$

$$=\frac{0.313\frac{r_2}{r_1}X_1}{2A_2}+\frac{0.313X_1}{A_1}$$

$$=X_1\left(\frac{0.313r_2}{2A_2r_1}+\frac{0.313}{A_1}\right)=X_1\beta$$

$$=\frac{0.313X_1}{A'}=\beta X_1$$

$$\therefore A'=\frac{0.313}{\beta}$$

where

$$\beta=\left(\frac{0.313r_2}{2A_2r_1}+\frac{0.313}{A_1}\right) \tag{23}$$

The value obtained for A' is the equivalent pole-face area and may be used in the calculations for force and reluctance. Thus

$$\mathcal{R}_g=\mathcal{R}_o+\mathcal{R}_a+\frac{0.313\,X}{A'} \tag{24}$$

$$\mathcal{R}_t=\frac{\mathcal{R}_l\mathcal{R}_g}{\mathcal{R}_l+\mathcal{R}_g} \tag{25}$$

In the expression for the value of flux (ϕ_{go}) necessary to start motion A' must be used in place of A.

$$\phi_{go}=\left(\frac{F_oA'}{0.0139}\right)^{1/2} \tag{26}$$

If these conditions of equivalence are met, the problem may be handled as the 2-leg case.

11. The solution for the 2-leg and E-core cases is completed by the reiterative calculation of the following equations in the sequence shown. The derivation of the current equation is given in the Appendix.

$$i_n=\left\{\frac{L_n-2}{L_{n-1}}\,i_{n-1}{}^2+\frac{L_oE^2}{L_{n-1}R^2}\times\right.$$

$$\left[\left(1-\epsilon^{-\frac{Rt_n}{L_o}}\right)^2-\left(1-\epsilon^{-\frac{Rt_{n-1}}{L_o}}\right)^2\right]-0.113\times$$

$$\left.\left[\frac{M'(V_{n-1}{}^2-V_{n-2}{}^2)+K(S_{n-1}{}^2-S_{n-2}{}^2)+(2F_o+P_{n-1})(S_{n-1}-S_{n-2})}{L_{n-1}}\right]\right\}^{1/2} \tag{27}$$

$$\phi_{gn}=\frac{Ni_n(10^{-8})}{\mathcal{R}_{gn-1}} \tag{28}$$

$$f_n=\frac{0.0139\phi_{gn}{}^2}{A'} \tag{29}$$

$$F_{An}=f_n-(F_o+P_{n-1}+KS_{n-1}) \tag{30}$$

$$V_n=\frac{\Delta t}{M'}\left(F_{A1}+F_{A2}+F_{A3}\ldots+\frac{F_{An}}{2}\right) \tag{31}$$

$$S_n=\Delta t\left(V_1+V_2+V_3\ldots+\frac{V_n}{2}\right) \tag{32}$$

$$P_n=f(S) \tag{33}$$

$$X_n=X_o-S_n \tag{34}$$

$$\mathcal{R}_{an}=\frac{C_2l_a}{a_ac_1-\phi_{gn}}\times10^{-8} \tag{35}$$

$$\mathcal{R}_{gn}=\mathcal{R}_{an}+\mathcal{R}_o+\frac{0.313X_{n-1}}{A'} \tag{36}$$

$$\mathcal{R}_{tn}=\frac{\mathcal{R}_l\mathcal{R}_{gn}}{\mathcal{R}_l+\mathcal{R}_{gn}} \tag{37}$$

$$L_n=\frac{N^2(10^{-8})}{\mathcal{R}_{tn}} \tag{38}$$

The problem is terminated when X=0.

3-LEG CASE

The solution for the 3-leg case is slightly more complicated than for the 2-leg and E-core cases, because the flux divides between the two outer legs in proportion to the sum of the reluctances in the parallel flux paths. By means of approximations, a solution can be obtained. Additional work is to be done to determine the degree of accuracy and possibly modify this method.

1. The mechanical forces are equated to act at the center of the main working gap as in the 2-leg case. Items 1–3, under "Method of Solution" also hold for the 3-leg case.

2. In the calculation of ϕ_{go}, the effect of the force generated at the outer leg will be small and may be ignored, therefore:

$$\phi_{go}=\left(\frac{F_o\,A'}{0.0139}\right)^{1/2} \tag{39}$$

At some point in the dynamic portion of the solution the effect of the outer working gap may become quite large and it must be taken into account in the second stage of the solution.

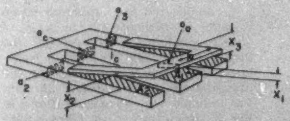


Fig. 7. Parameters of E-core magnet structure

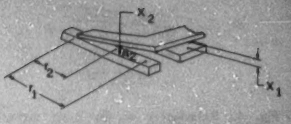


Fig. 8. Air-gap reluctance of E-core magnet

3. Once the physical parameters of the magnet have been established, it is necessary to calculate the total reluctance of the magnetic circuit as indicated in Fig. 9. The approximate equivalent circuit is shown in Fig. 10.

The following restrictions apply to the initial conditions.

1. At the time when motion starts (t_o) the armature is operating below the knee of the magnetization curve.

2. The flux in the yoke is distributed proportionately to the assumed air gap at the hinge end and the gap at the outer leg.

3. The reluctance of both legs of the magnet are assumed constant at the value calculated for the leg at the hinge end. The restrictions in this case are similar to those imposed in the 2-leg case.

4. Before the initial reluctance may be determined, values for ϕ_{10} and ϕ_{20} must be calculated. The following relationship can be shown to be a good approximation.

$$\frac{\phi_2}{\phi_1}=\frac{\frac{0.313l_h}{a_H}}{\frac{0.313(2X)}{a_2}}\text{ assuming }r_2=2r_1$$

$$\phi_2=\frac{a_2}{a_H}\left(\frac{l_H}{2X}\right)\phi_1 \tag{40}$$

Therefore, it is assumed that the flux divides proportionate to the length of the hinge and outer gap. By definition the outer gap cannot be zero so X must be limited to the value l_H.

$$\phi_{go}=\phi_{10}+\phi_{20}$$

$$=\phi_{10}+\frac{a_2}{a_H}\left(\frac{l_H}{2X_o+l_H}\right)\phi_{10}$$

$$=\phi_{10}\left[1+\frac{a_2}{a_H}\left(\frac{l_H}{2X_o+l_H}\right)\right]$$

$$K_{84}=\frac{a_2}{a_H} \tag{41}$$

$$\phi_{10}=\frac{\phi_{go}}{1+K_{84}\left(\frac{l_H}{2X_o+l_H}\right)}$$

$$\phi_{20}=\phi_{go}-\phi_{10}$$

$$\mathcal{R}_{a10}=\frac{C_2l_{a1}(10^{-8})}{C_1a_{a1}-\phi_{10}} \tag{42}$$

$$\mathcal{R}_{a20}=\frac{C_2l_{a2}(10^{-8})}{C_1a_{a2}-\phi_{20}} \tag{43}$$

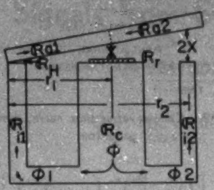


Fig. 9. Reluctances of the 3-leg magnet

$\mathcal{R}_{a1}$ = reluctance of armature (inner leg)
$\mathcal{R}_{a2}$ = reluctance of armature (outer leg)
$\mathcal{R}_{c}$ = reluctance of core
$\mathcal{R}_{i1}$ = reluctance of yoke (inner leg)
$\mathcal{R}_{i2}$ = reluctance of yoke (outer leg)
$\mathcal{R}_{H}$ = reluctance of hinge

$$\mathcal{R}_{i1}=\frac{C_2l_{i1}(10^{-8})}{C_1a_{i1}-\phi_{10}} \qquad (44)$$

$$\mathcal{R}_{i2}=\frac{C_2l_{i2}(10^{-8})}{C_1a_{i2}-\phi_{20}} \qquad (45)$$

$$\mathcal{R}_{c}=\frac{C_2l_{c}(10^{-8})}{C_1a_{c}-\phi_{go}} \qquad (46)$$

$$\mathcal{R}_{r}=\frac{0.313l_r}{a_r} \qquad (47)$$

$$\mathcal{R}_{H}=\frac{0.313l_H}{a_H} \qquad (48)$$

$\mathcal{R}_l$ is now solved.[1]

$$\mathcal{R}_{go}=\mathcal{R}_r+\mathcal{R}_c+\frac{0.313X_o}{A_1}+\left[\frac{(\mathcal{R}_{a10}+\mathcal{R}_{i1}+\mathcal{R}_H)\left(\mathcal{R}_{a20}+\mathcal{R}_{i20}+\frac{0.626X_o}{A_2}\right)}{\mathcal{R}_{a10}+\mathcal{R}_{i1}+\mathcal{R}_H+\mathcal{R}_{a20}+\mathcal{R}_{i20}+\frac{0.626X_o}{A_2})}\right] \qquad (49)$$

$$\mathcal{R}_{to}=\frac{\mathcal{R}_l\mathcal{R}_{go}}{\mathcal{R}_l+\mathcal{R}_{go}} \qquad (50)$$

$$L_o=\frac{N^2(10^{-8})}{\mathcal{R}_{to}} \qquad (51)$$

$$i_o=\frac{\mathcal{R}_{go}\phi_{go}(10^8)}{N} \qquad (52)$$

$$t_o=\frac{L}{R}\ln\left(\frac{E}{E-Ri_o}\right) \qquad (53)$$

The expression for P_n may be generated as shown in item 9 in the discussion of "Method of Solution."

The solution for the 3-leg case is completed by the reiterative calculation of the following equations in the sequence shown. The derivation of the current equation is given in the Appendix.

$$i_n=\left\{\frac{L_{n-2}}{L_{n-1}}\,i_{n-1}^2+\frac{L_oE^2}{L_{n-1}R^2}\times\right.$$

$$\left.\left[\left(1-\epsilon^{-\frac{Rt_n}{L_o}}\right)^2-\left(1-\epsilon^{-\frac{Rt_{n-1}}{L_o}}\right)^2\right]-0.113\times\left[\frac{M'(V_{n-1}^2-V_{n-2}^2)+K(S_{n-1}^2-S_{n-2}^2)+(2F_o+P_{n-1})(S_{n-1}-S_{n-2})}{L_{n-1}}\right]\right\}^{1/2} \qquad (54)$$

$$\phi_{gn}=\frac{Ni_n(10^{-8})}{\mathcal{R}_{gn-1}} \qquad (55)$$

$$\phi_{1n}=\frac{\phi_{gn}}{1+K_{34}\dfrac{l_H}{2X_n+l_H}} \qquad (56)$$

$$\phi_{2n}=\phi_{gn}-\phi_{1n} \qquad (57)$$

$$f_n=0.0139\left[\frac{\phi_{gn}^2}{A_1}+\frac{2\phi_{2n}^2}{A_2}\right] \qquad (58)$$

$$F_{An}=f_n-(F_o+P_{n-1}+KS_{n-1}) \qquad (59)$$

$$V_n=\frac{\Delta t}{M'}\left(F_{A1}+F_{A2}+F_{A3}\ldots+\frac{F_{An}}{2}\right) \qquad (60)$$

$$S_n=\Delta t\left(V_1+V_2+V_3\ldots+\frac{V_n}{2}\right) \qquad (61)$$

$$P_n=f(S_n) \qquad (62)$$

$$X_n=X_o-S_n \qquad (63)$$

$$\mathcal{R}_{a1n}=\frac{C_2l_{a1}(10^{-8})}{C_1a_{a1}-\phi_{1n}} \qquad (64)$$

$$\mathcal{R}_{a2n}=\frac{C_2l_{a2}}{C_1a_{a2}-\phi_{2n}}\times10^{-8} \qquad (65)$$

$$\mathcal{R}_{gn}=\mathcal{R}_r+\mathcal{R}_c+\frac{0.313X_n}{A}+\left[\frac{(\mathcal{R}_{a1n}+\mathcal{R}_{i1}+\mathcal{R}_H)\left(\mathcal{R}_{a2n}+\mathcal{R}_{i2}+\frac{0.626X_n}{A_2}\right)}{\mathcal{R}_{a1n}+\mathcal{R}_{i1}+\mathcal{R}_H+\mathcal{R}_{a2n}+\frac{0.626X_n}{A_2}+\mathcal{R}_{i2}}\right] \qquad (66)$$

$$\mathcal{R}_{tn}=\frac{\mathcal{R}_l\mathcal{R}_{gn}}{\mathcal{R}_l+\mathcal{R}_{gn}} \qquad (67)$$

$$L_n=\frac{N^2(10^{-8})}{\mathcal{R}_{tn}} \qquad (68)$$

The problem is terminated when $X=0$.

Summary

The results of the computations give a good approximation within the stated limits. Several sample problems have been computed and the results checked by comparison with measurements of operating models. The method has also been applied to improve the design of relay magnet structures, with excellent results.

For the first stage of the operate solution it is not necessary that the solution be obtained manually. The computer may be programmed to start at $t_o=0$.

The utility of this method is the ability to run a large number of solutions with a minimum human effort. As the parameters of the magnet are varied, the approach to the optimum operating characteristics may be analyzed and additional direction given to the design. For example, the International Business Machines Corporation Card Programmed Calculator will provide a solution in approximately 15 minutes for a magnet which requires 4 milliseconds to seal. In an 8-hour day, 32 magnet designs may be evaluated.

It should be noted that the effects of eddy currents do not appear in the solution. In some types of magnets, the eddy currents have a large effect on the operate characteristics. In these cases it would be noted that the measured current would be substantially different than the computed current even though the displacement, velocity, etc., agree quite well. It may be assumed that the greater part of this difference is due to eddy currents.

Some initial work has been started to include the effects of eddy currents, but it is not sufficiently advanced to be included in this report.

It may also be pointed out that some of the assumptions were based on rather meager data. Further investigation is necessary in this field, but it is felt that in its present state, the information in this report is a useful aid in magnet design.

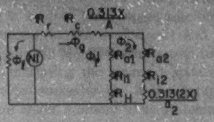


Fig. 10. Equivalent circuit of the 3-leg magnet

$\frac{0.313X}{A}$ = reluctance of center working gap

$\frac{0.313(2X)}{A_2}$ = reluctance of outer leg working gap

$\mathcal{R}_r$ = reluctance of residual
$\mathcal{R}_l$ = leakage reluctance

Appendix. Derivation of Current Equation

It has been shown that the equation for inductance is as follows:

$$L=\frac{N^2(10^{-8})}{\mathcal{R}_t}$$

In an electromagnet with the armature in an open position, it may be shown that a large portion of the total reluctance resides in the air gaps. For the moderate values

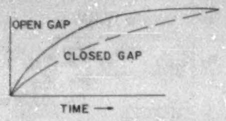


Fig. 11. Current versus time curve, static condition

of H encountered in the types of magnets under discussion the reluctance may be assumed constant for the open-air-gap case. Likewise then, the inductance may be assumed constant.

If the magnet armature is blocked closed, the foregoing is no longer the case and the reluctance is greatly reduced at low values of flux and becomes a function of the magnitude of the flux. Generally it may be assumed that the average inductance is much greater in this case.

The dynamic case, wherein the armature is allowed to operate as the current is applied, must lie somewhere between the foregoing examples.

The current rise for the case of constant inductance may be derived from the following expression:

$$L\frac{di}{dt}+Ri=E$$

thus

$$i=\frac{E}{R}\left(1-e^{-\frac{Rt}{L}}\right)$$

If the average closed gap inductance is considered, this expression will hold. The current will rise in the manner indicated in Fig. 11.

In the dynamic case the current will rise as shown in Fig. 12.

In the dynamic case the fact that L is a variable must be considered; therefore

$$L\frac{di}{dt}+Ri=E \quad L=f(\phi,i)\pm\text{constant}$$

This equation does not lend itself to solution by rigorous means, but if the energy relationships involved in the operation of the magnet are analyzed, a method of solution is suggested.

When voltage is applied to the coil of an electromagnet an energy transfer is involved. A portion of the energy is stored in the magnetic field and a portion is dissipated as heat in the I^2R loss in the magnet coil. Of the energy which has gone into the generation of the field, a portion is irreversibly lost in the hysteresis effects in the iron and in the generation of eddy currents. The remainder of the field energy may be used to do mechanical work.

If the energy transferred to the magnet with the armature blocked open is considered, total input energy

$$\epsilon=\int e\,i\,dt$$

$$=\int\left(L\frac{di}{dt}+Ri\right)idt$$

$$=\frac{1}{2}Li^2+\int Ri^2dt$$

the energy dissipated

$$\epsilon_D=\int Ri^2dt$$

the field energy

$$\epsilon_F=\frac{1}{2}Li^2$$

If no motion of the armature occurs, the inductance is assumed constant. Therefore

$$i=\frac{E}{R}\left(1-e^{-\frac{Rt}{L}}\right)$$

$$\epsilon_F=\frac{1}{2}Li^2$$

$$=\frac{1}{2}L\left(\frac{E^2}{R^2}\left(1-e^{-\frac{Rt}{L}}\right)^2\right)$$

$$=\frac{E^2L}{2R^2}\left[\left(1-e^{-\frac{Rt}{L}}\right)^2\right]$$

where

L = henrys
i = amperes
E = joules

If the armature is allowed to move as a result of the magnetic pull the mechanical energies involved are as follows:

$$\epsilon_M=\int MVdv+\int KSds+\int\left(F_o+\frac{P}{2}\right)ds$$

$$=\frac{MV^2}{2}+\frac{KS^2}{2}+\frac{(2F_o+P)}{2}S \quad \epsilon m=\text{in}-\text{lbs}$$

where

$\frac{MV^2}{2}$ = kinetic energy

$\frac{KS^2}{2}$ = energy stored in the return spring

therefore

$\frac{(2F_o+P)}{2}S$ = energy expended in the mechanical system

$$\epsilon_M=0.113\left[\frac{MV^2+KS^2+(2F_o+P)S}{2}\right]$$

The apparent change in field energy is equal to the total field energy minus mechanical energy

$$\Delta\epsilon_F=\epsilon_F-\epsilon_M$$

This may be written in difference form as follows:

$$\left.\frac{Li^2}{2}\right]_{t_{n-1}}^{t_n}=\left\{\left.\frac{L_oE^2}{2R^2}\left(1-\epsilon^{-\frac{Rt}{L_o}}\right)^2\right]_{t_{n-1}}^{t_n}-0.113\times\left[\frac{MV^2+KS^2+(2F_o+P)S}{2}\right]_{t_{n-1}}^{t_n}\right\}$$

$$\frac{L_{n-1}i_n^2-L_{n-2}i_{n-1}^2}{2}=\left\{\frac{L_oE^2}{2R^2}\times\right.$$

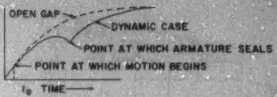


Fig. 12. Current versus time curve, dynamic condition

$$\left[\left(1-\epsilon^{-\frac{Rt_n}{L_o}}\right)^2-\left(1-\epsilon^{-\frac{Rt_{n-1}}{L_o}}\right)^2\right]-$$

$$\left.0.113\left[\frac{M(V_n^2-V_{n-1}^2)+K(S_n^2-S_{n-1}^2)+(2F_o+P_n)(S_n-S_{n-1})}{2}\right]\right\}$$

$$L_{n-1}i_n^2=\left\{L_{n-2}i_{n-1}^2+\frac{L_oE^2}{R^2}\times\right.$$

$$\left[\left(1-\epsilon^{-\frac{Rt_n}{L_o}}\right)^2-\left(1-\epsilon^{-\frac{Rt_{n-1}}{L_o}}\right)^2\right]-$$

$$\left.0.113[M(V_n^2-V_{n-1}^2)+K(S_n^2-S_{n-1}^2)+(2F_o+P_n)(S_n-S_{n-1})]\right\}$$

$$i_n^2=\left\{\frac{L_{n-2}}{L_{n-1}}i_{n-1}^2+\frac{L_oE^2}{L_{n-1}R_2}\times\right.$$

$$\left[\left(1-\epsilon^{-\frac{Rt_n}{L_o}}\right)^2-\left(1-\epsilon^{-\frac{Rt_{n-1}}{L_o}}\right)^2\right]-$$

$$\left.0.113\left[\frac{M(V_n^2-V_{n-1}^2)+K(S_n^2-S_{n-1}^2)+(2F_o+P_n)(S_n-S_{n-1})}{L_{n-1}}\right]\right\}$$

$$i_n=\left\{\frac{L_{n-2}}{L_{n-1}}i_{n-1}^2+\frac{L_oE^2}{L_{n-1}R^2}\times\right.$$

$$\left[\left(1-\epsilon^{-\frac{Rt_n}{L_o}}\right)^2-\left(1-\epsilon^{-\frac{Rt_{n-1}}{L_o}}\right)^2\right]-$$

$$\left.0.113\left[\frac{M(V_n^2-V_{n-1}^2)+K(S_n^2-S_{n-1}^2)+(2F_o+P_n)(S_n-S_{n-1})}{L_{n-1}}\right]\right\}^{1/2}$$

References

1. DESIGN OF RELAYS. *Bell System Technical Journal*, New York, N. Y., vol. 33, Jan. 1954.

2. ELECTRIC AND MAGNETIC FIELDS (book), S. S. Attwood. John Wiley & Sons, Inc., New York, N. Y., third edition, 1949.

3. ELECTRO MAGNETICS (book), J. D. Kraus. McGraw-Hill Book Company, Inc., New York, N. Y., 1933.

4. ELECTRICAL ENGINEERING (book), G. V. Mueller. McGraw-Hill Book Company, Inc., second edition, 1948.

5. ELECTRO-MAGNETIC DEVICES (book), H. C. Roters. John Wiley & Sons, Inc., New York, N. Y., first edition, Dec. 1951.

6. MAGNETIC CIRCUITS AND TRANSFORMERS (book), Electrical Engineering Staff of Massachusetts Institute of Technology. John Wiley & Sons, Inc., July 1950.

Signals from Switched Ferroelectric Memory Capacitors

C. F. PULVARI
NONMEMBER AIEE

G. E. McDUFFIE, JR.
NONMEMBER AIEE

CONSIDERABLE information has been published on the properties of ferroelectric materials[1-4] and their use in digital memory devices has been described.[5,6] It is the purpose of this paper to discuss the output signals from switched and unswitched ferroelectric storage capacitors, to treat the storage capacitor as a circuit element by the introduction of an equivalent circuit, and to present methods and several circuits by which the inherent signal-to-noise ratio can be improved.

The Bistable Ferroelectric Storage Element

The electrostatic hysteresis loop describes the macroscopic behavior of ferroelectrics and shows that reversibility of (remanent) polarization by an applied field is the real criterion of ferroelectric behavior. Hysteresis loops have been obtained for polycrystalline ceramics as well as for single crystals: loops for an aggregation of single crystals in a polycrystalline matrix can also be obtained. Hysteresis loops are characteristic of the material.

The hysteresis loop of a perfect single crystal is practically a rectangle. Fig. 1 shows a hysteresis curve obtained from a c-domain crystal of barium titanate at a frequency of 20 kc. After application and removal of each polarizing field, a certain remanent polarization P_r is obtained. Maximum obtainable value of P_r, or the remanent polarization, is obtained from a saturated hysteresis loop. Remanent polarization binds on the capacitor electrodes certain quantities of charge Q. If it is assumed that the remanent polarization is partially or completely changed, charge Q, corresponding to P_r, is released and may be used to obtain a signal. When a change in polarization occurs, the quantity of bound charges is altered such that during an increase in polarization, the dielectric seems to absorb charges and during a decrease in polarization, the dielectric appears to liberate charges. Thus, a ferroelectric dielectric material resembles, in some respects, a storage device; charges (electric energy) can be stored by polarizing the ferroelectric dielectric of a storage capacitor but, unlike a linear capacitor, the bound charges cannot be removed by short-circuiting the storage capacitor.

For memory applications a nearly rectangular hysteresis loop is desired; two important quantities defined by this loop are P_r, the remanent polarization, and E_c the coercive force field. E_c is defined as the field at which the polarization vanishes. In plotting P instead of D, dielectric displacement in air caused by applied field E is neglected since its value is approximately only 0.1 per cent of the remanent polarization in the case of a nearly rectangular loop.

The Bistable Capacitor as a Circuit Element

BASIC SWITCHING CIRCUIT

Fig. 2 shows the basic circuit of a single ferroelectric memory flip-flop. C_s denotes the storage capacitor; the load, which may take various forms, is symbolized by the box. R_0 is the internal resistance of the source. In order to store digital information in the storage element, a reference frame must be chosen. Assume that the $+P_r$ point on the hysteresis loop is the normal rest state of the element, and that it corresponds to a binary zero (0). A negative voltage of sufficient amplitude will switch the element from the $+P_r$ state to the $-P_r$ state; let the $-P_r$ state correspond to a binary one (1). Thus the memory elements of an array of condensers normally store binary 0's; digital words are written into the array by the application of negative write-pulses corresponding to the digital 1's present in the word.

If capacitor C_s stores a binary 1, the sensing or readout pulse causes a switching transient in the load circuit and a large signal will appear on the output terminals, indicating the binary 1 state of the selected capacitor C_s; if the storage capacitor stores a binary 0, no switching will occur and only a small elastic transient will be caused. The ratio of switched to nonswitched transient is called signal-to-noise ratio; the numerical value of this ratio will depend on the type of load used in Fig. 2 and upon the time at which the two signals are measured.

The main point in this consideration is that the storage capacitor C_s can be regarded as a transient-current source, since this is determined only by the properties of the storage capacitor such as P_r, area and thickness of the element, and its crystalline structure, provided that the internal resistance of the switching-pulse source R_0 is small.

DYNAMIC SWITCHING RESISTANCE CURVES

It has been shown that, during switching, the storage capacitor behaves like a resistance.[7] Fig. 3 shows a family of curves obtained by plotting the maximum switched current as a function of the input pulse voltage using different resistive loads. The current wave form for resistive loads is shown in Fig. 4. Dynamic switching resistance R_D is defined as the reciprocal of the slope of these curves; it is dynamic in the sense that the storage capacitor is not a steady-state device and therefore can only be measured by variations of the input voltage. From these curves the peak power supplied to the storage capacitor during switching can be determined for various resistive loads by taking the product of the maximum current and the input pulse voltage.

EQUIVALENT CIRCUIT

The family of dynamic switching resistance curves also indicates the type of equivalent circuit most suited to the storage capacitor. For a given value of input pulse voltage a variation in resistive load of approximately 1000 to 1 produces a current change of about 2 to 1, indicating an equivalent source of the current type. Once this family of curves is determined for a particular storage capacitor, an equivalent circuit can be established.

For most switching applications a convenient equivalent circuit for the bistable storage capacitor is a current generator with a resistance in parallel. When a step function is applied, the current

---

Paper 56-755, recommended by the AIEE Solid State Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted September 1, 1955; made available for printing May 3, 1956.

C. F. PULVARI AND G. E. McDUFFIE, JR., are with The Catholic University of America, Washington, D. C.

The work reported had the financial support of the U. S. Air Force, Aeronautical Research Laboratory, Wright Air Development Center, Dayton, Ohio, under *Contract AF No. 18(600)-106.*

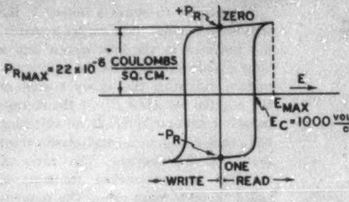


Fig. 1. Hysteresis loop of c-domain barium titanate

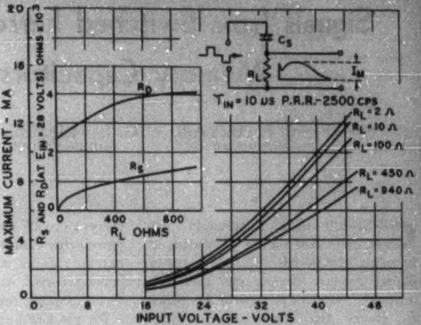


Fig. 3 (right). Dynamic switching resistance curves

generator produces a current transient as shown in Fig. 4; the elastic component of capacity C_0 can normally be neglected. From these considerations, the equivalent circuit shown in Fig. 5 can be derived where $i(t)$ designates the current generator producing a current transient; the shunt resistance, R_s, can be computed from the family of dynamic switching resistance curves, for a given applied input voltage, from the relation

$$R_s = \frac{R_L}{\frac{I_{MO}}{I_M} - 1} \tag{1}$$

where R_L is the load resistance, I_{MO} is the maximum current for $R_L = 0$, and I_M is the maximum current for the value of R_L under consideration. The inset curve in Fig. 3 shows the variation of R_D and R_S with change in load resistance. Although the ohmic value of R_S is relatively small, the ratio R_S/R_L in the range of interest is sufficiently large to justify the current source equivalent.

Different types of impedance may be applied as a load, such as a resistance, capacitance, inductance, or a nonlinear resistance of any type. Examples of such loads are resistors, capacitors, pulse transformers, and diodes or transistors. Load circuits may be formed by combining different impedances as desired. The character of the output signal depends on the type of load chosen; see Figs. 4 and 6. The load impedance may be a part of the readout circuit; different types of circuit are discussed in the following.

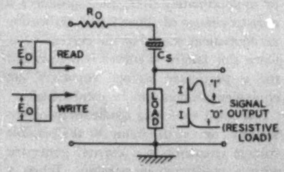


Fig. 2. Basic storage circuit

If an electronic component is to be used in circuitry for data handling or processing devices, a thorough knowledge of its switching properties is of primary interest. To facilitate mathematical treatment of related problems, a mathematical expression was developed for the switching transient.

Referring to Fig. 4, the current due to the switched charge can be approximated by

$$i(t) = \alpha I_M t e^{-\frac{t^2}{4a}} \tag{2}$$

where the maximum switching current I_M and the constant a depend on the applied switching field, and α is a constant, whose value depends on P_r and the area of the ferroelectric element; the fitting constants a and α are discussed in detail in reference 8. Equation 2 neglects the initial spike of current due to the linear component of capacity.

Illustration

As an application of this function, the output voltage due to the switching current across a parallel resistance-capacitance load has been calculated. Fig. 7(A) shows the circuits under consideration. $e(t)$ is the actual voltage developed across R_LC_L during switching; it consists of the sum of the voltage developed by the switching current $i(t)$ and the linear circuit voltage drop. A typical wave form of $e(t)$ is shown in Fig. 6. For purposes of calculation, the equivalent circuit shown in Fig. 7(B) is used; $e'(t)$ is the component of output voltage due to the switching current alone. In this example R_s, the internal shunting resistance across the current generator may be neglected if the time constant R_LC_L is chosen such that the duration of the output voltage $e'(t)$ is much larger than the duration of the current transient $i(t)$. In the case of other load impedances in which the output transient is comparable in duration to the current transient, R_s may not be neglected. Assuming the time constant R_LC_L to be such that R_s may be neglected, the equation describing behavior of the equivalent circuit is

$$\alpha I_M t e^{-\frac{t^2}{4a}} = \left(\frac{1}{R_L} + C_L \frac{d}{dt}\right) e'(t) \tag{3}$$

By considering numerical value of α, I_M and a, and neglecting numerically small terms in the exact solution of equation 3, an approximate solution can be shown to be

$$e'(t) \cong \frac{2a\alpha I_M}{C_L}\left(e^{-\frac{t}{R_LC_L}} - e^{-\frac{t^2}{4a}}\right) \tag{4}$$

Fig. 7 shows experimental and calculated values of $e'(t)$ for $C_L = 5{,}000$ micromicrofarads and $R_L = 4{,}700$ ohms. R_L was chosen at this relatively high value in order to increase the time scale so as to provide a better check on the calculated values. Typical values of the constants for small storage capacitors ($A = 10^{-2}\text{cm}^2$) are $a = 1.54 \times 10^{-12}\text{sec}^2$, $\alpha = 2.1 \times 10^6\ \text{sec}^{-1}$ and $I_m = 3 \times 10^{-3}$ amperes.

Reading Modes

To recover stored information from the ferroelectric memory element, three general methods of sensing the state of the element are possible:

1. Destructive readout.
2. Nondestructive readout.
3. Multiple readout.

Vol.
Pt. 1
1956

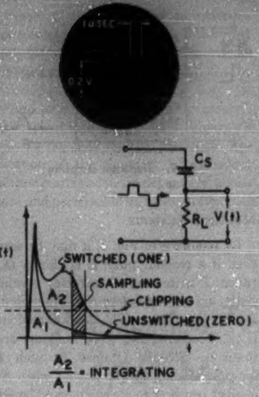


Fig. 4. Transients of a resistive load: $R_L = 220$ ohms

Destructive Readout

In destructive readout, a positive voltage is applied of sufficient amplitude to switch those memory elements in the $-P_r$ state to the $+P_r$ state. If a binary 1 is stored, switching occurs, giving rise to an output signal; if binary 0 is stored, no switching occurs and little or no output signal is obtained. Ideally, the detection of a binary 1 or 0 is a simple matter; however, when one element of an array of storage elements is sensed, the ratio of output signal for a binary 1 to the output signal for a binary 0 is well below the ideal case since the shunting capacity of the other element is appreciable.

Shunting capacity as well as series load impedance introduces a fraction of the sensing pulse, $e(t)-e'(t)$, into the output signal. The purpose of the following readout systems is to decrease or eliminate the effect of the sensing pulse and shunt capacity on the 1-to-0 output signal ratio.

Fig. 4 shows the output currents of a single storage element corresponding to a binary 1 and 0. The ratio of the two outputs can be artificially increased by decreasing or eliminating the sensing pulse through clipping, integrating, or sampling.

When clipping is used for readout, a reference level is established slightly above the amplitude of the output signal due to a binary 0. Fig. 8 shows a simple diode-clipping circuit. Since clipping may operate on a voltage produced during the actual sensing pulses, access time is very rapid.

Integrating readout compares the actual switched-to-nonswitched-charge ratio, represented by the areas A_2/A_1 in Fig. 4, rather that the ratio of current amplitudes. In this mode, a parallel resistance capacitance combination is used as the load impedance. The capacity integrates the charge while the parallel resistance acts as a discharge path. Fig. 6 shows the output signals for switched and nonswitched operation obtained across a resistance capacitance load. It may be noted that the signal does not fall to zero at the end of the pulse; thus, for complete integration, access time is inherently increased. The parallel resistance, however, may be decreased until the duration of the output signal is the same as, or less than the duration of the sensing pulse; under these conditions complete integration is not obtained. The output from the integrating circuit may be fed into a clipping circuit further to increase the switched to nonswitched ratio.

Capacity of the integrating capacitor should be sufficiently high to provide complete integration, yet it should be low enough to produce a workable signal level. It has been shown[8] that an optimum value exists which is given by

$$C_L(\text{opt}) = \frac{2AP_r}{\delta E_c} \tag{5}$$

where A is the area and δ the thickness of the element.

Sampling of the output signal may be accomplished by time sampling or by frequency sampling.

Time sampling is performed by time gating a portion of the output signal. This mode may be used on either the current wave form, as shown in Fig. 4, or on the charge wave form, as shown in Fig. 6. For ceramic storage elements of low sensitivity, it was found that time sampling of the output signal from an integrating load, during a short period immediately after the end of the sensing pulse, had certain advantages. Fig. 9 shows a simple transistor gate circuit which allows an output signal to be developed at its terminals only after the sensing pulse has dropped to zero.

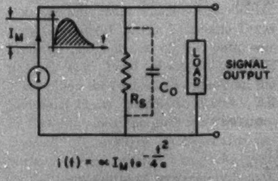


Fig. 5. Equivalent circuit

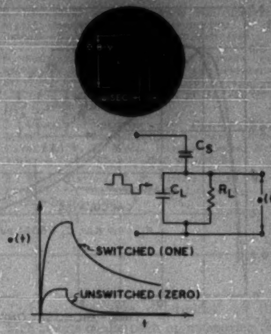


Fig. 6. Transients on a resistance-capacitance load: $R_L = 4700$ ohms, $C_L = 5{,}000$ microfarads

Frequency sampling separates the sensing pulse, developed across the load, from the integrated switched current by means of a suitable filter; it is particularly applicable when very short sensing pulses are used. This method utilizes a load whose time constant is long compared to the duration of the sensing pulse. Fig. 10 shows oscillograms of filtered and unfiltered output signals for 0.1-microsecond sensing pulses. A pulse amplifier with a relatively low-frequency cut-off (approximately 1 megacycle) was used as a filter.

Readout modes as described have been tested under varying conditions and proved satisfactory. Their characteristics and application may be summarized as follows:

1. Clipping: applicable to any type of load impedance; fast access time; requires a basic high switched-to-nonswitched ratio.
2. Integrating: capacitive-resistive load; slow access time; gives high basic switched-to-nonswitched ratio.
3. Time sampling: applicable to most types of load; medium access time; permits selection of optimum switched-to-nonswitched ratio; well suited for elements of low sensitivity.
4. Frequency sampling: applicable to most loads; medium access time; well suited for high-speed or multiple reading.

The different readout methods described may be readily combined according to need. As a practical example assume that a low-level switching transient is to be examined or utilized such as a signal from a minor loop or a signal from a material having low saturation polarization. After choosing a type of load

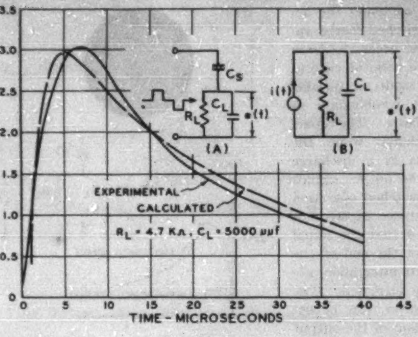


Fig. 7. Signal voltage, e'(t)

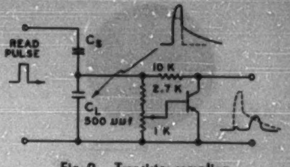


Fig. 9. Transistor sampling

suited to the desired access time, readout technique may be as follows: (1) The low level signal developed on the load is clipped to remove the sensing pulse. (2) The residual switching transient is amplified. (3) Time sampling is performed on the amplified transient such as to produce a suitable signal.

When a destructive readout mode is used and it is desired that the stored information be retained, a restoring type of operation is used. Such a readout-restore mode can be accomplished by using a feedback loop connected between output terminals of the readout circuit and the writing circuit. When a binary 1 is sensed in a memory element by a positive readout pulse, the writing circuitry is triggered, by which a negative write pulse is applied to the element thus restoring it to its original state. In this way, information in the storage elements can be "read" any desired number of times.

Nondestructive Readout

1. *Ultrasonic Mode.* Since polarized ferroelectric capacitors are inherently piezoelectric, it is possible to apply an ultrasonic field to a storage capacitor which in turn would deliver an a-c signal with the frequency of the applied ultrasonic vibrations. A stored 0 may be distinguished from a stored 1 by using a phase-sensitive output circuit connected to the load circuit, whereby a 180-degree phase difference will be observed depending on the state of the bistable capacitor. This type of readout was extensively used at the beginning and found to be practical. It was also used for detecting polarization of fine ferroelectric particles imbedded in an insulating binder; it served well for this purpose.

2. *High-Frequency Readout Modes.* Consideration was given to the use of two different high-frequency signals for energizing the rows and columns of a storage matrix but the method was not tested. In this mode nonlinearity of the storage capacitor can be used for producing a beat frequency. After rectifying, a phase-sensitive output circuit may again be used with advantage. Polarization states of the storage capacitor appear on the output with a phase difference of 180 degrees.

3. *Cross-Field Method.* In this method, two pairs of electrodes, mutually perpendicular, are applied on a single storage capacitor. One pair, applied on planes perpendicular to the *c*-axis, serve for writing, and the other, on planes perpendicular to the *a*-axis, for readout. Depending on the polarization state of the storage capacitor, the signal appears in the readout-lead with a phase difference of 180 degrees. This method was not tested.

Multiple Readout

By reference to Fig. 1, it may be seen that, if a readout pulse *E* is chosen between *E*-write and *E*-coercive, several pulses are needed before P_r pertaining to *E*-write is read out. The closer *E*-readout is to *E*-coercive, the more readouts are possible. Signals obtained in such a multiple readout will be smaller when *E*-readout is nearer to *E*-coercive and larger as their difference increases. This mode was satisfactorily tested; it may be useful in some special cases.

Although not much work has been performed on nondestructive readout systems, it was realized early in this work that in a later development it would be desirable to develop readout modes whereby stored information may be used as many times as desired.

Conclusions

In circuit applications of a bistable ferroelectric memory capacitor, it is important to determine the signal ratios for swithced and nonswitched conditions for a particular load circuit. The signal ratio represents the inherent signal-to-noise ratio for a memory capacitor. Since the attainable signal-to-noise ratio is one of the limiting factors in application, it was necessarily important that the inherent signal-to-noise ratio be improved by using additional readout circuitries.

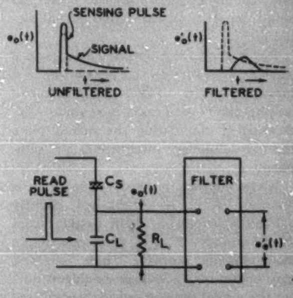


Fig. 10. Filtering circuit

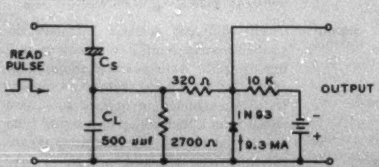


Fig. 8. Diode-clipping circuit

Inherent nonlinearity of ferroelectric storage elements prevents straightforward analysis when they are used in circuit application. It has been shown that a ferroelectric storage element may be treated as a transient current generator. The type of load used to detect binary 1's and 0's depends on the particular application of the element. The readout method employed again depends on access time and permissible signal-to-noise ratio.

References

1. THEORETICAL MODEL FOR BARIUM TITANATE, W. P. Mason, B. T. Matthias. *Physical Review*, New York, N. Y., vol. 79, Dec. 1948, pp. 1622–36.

2. THEORY OF FERROELECTRICS, A. F. Devonshire. *Advances in Physics*, London, England, vol. 3, Apr. 1954, pp. 86–130.

3. THE LORENTZ CORRECTION IN BARIUM TITANATE, J. C. Slater. *Physical Review*, New York, N. Y., vol. 78, June 1950, pp. 748–61.

4. FERROELECTRICITY, DOMAIN STRUCTURE AND PHASE TRANSITIONS OF BARIUM TITANATE, A. von Hipple. *Reviews of Modern Physics*, New York N. Y., vol. 22, July 1950, pp. 221–37.

5. APPARATUS AND A METHOD FOR ELECTROSTATIC RECORDING AND REPRODUCING, C. F. Pulvari. *U. S. Patent No. 145,361*, Feb. 1950.

6. FERROELECTRIC MATERIALS AS STORAGE ELEMENTS FOR DIGITAL COMPUTERS AND SWITCHING SYSTEMS, J. R. Anderson. *AIEE Transactions (Electrical Engineering)*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 395–401.

7. DOMAIN FORMATION AND DOMAIN WALL MOTION IN FERROELECTRIC $BaTiO_3$, SINGLE CRYSTALS, W. T. Merz. *Physical Review*, New York, N. Y., vol. 95, Aug. 1954, pp. 690–98.

8. FINAL REPORT, C. F. Pulvari. *USAF Contract No. 18(600)-106*, Wright Air Development Center, Dayton, Ohio, Apr. 1955.

Propagation Test on 955.5 Mc, 1,965 Mc, and 6,730 Mc

H. R. MATHWICH
NONMEMBER AIEE

E. D. NUTTALL
MEMBER AIEE

J. E. PITMAN
AFFILIATE MEMBER AIEE

A. M. RANDOLPH
ASSOCIATE MEMBER AIEE

THE rapidly expanding need for communications felt by United Gas Pipe Line Company indicated an examination of the new communication medium, microwave relay. Work by Bullington,[1] Crawford and Jakes,[2] Durkee,[3] Millar and Byam,[4] Thompson,[5] and others, provided valuable information for system planning and operation in most areas of the United States. In the Gulf South area, locale of the company's operations, no such information was available. Rumors, difficult to prove or disprove, were current concerning the propagation difficulties encountered by early systems in the area. In this almost subtropical section of the country, humidity variations were known to be high. Temperature inversion also had to be considered. Previous observations of very-high-frequency systems and a test at 960 mc (megacycles) indicated the possibility of unusual conditions at the higher frequencies of 2,000 and 6,700 mc.

Investigation by the Research Department of United Gas revealed the interest of the microwave manufacturers in the propagation problems of the area. A test was designed to gather data on three of the frequency bands available to pipeline companies. It was planned to operate circuits at 955.5, 1,965, and 6,730 mc over the same path for a period long enough to provide statistical data. United Gas, as the prime contractor, arranged for Philco Corporation to furnish certain equipment and services for the 6,730-mc frequencies. By separate arrangements with The Radio Corporation of America (RCA), equipment and services relating to the 955.5- and the 1,965-mc frequency were made available. The results of the ensuing test are reported in this paper. Its purpose is not to judge the relative merits of the three frequency bands involved, but rather to make the accumulated data available to the literature. It must be constantly borne in mind in the interpretation of the data presented that the tower heights and resulting path clearances were not considered optimum for any of the three frequencies being tested.

Description of Test

The test circuits were set up between the company's research laboratory at Shreveport, La., and a compressor station at Latex, Tex., a distance of 20.25 miles. A strip map of the path is shown in Fig. 1 and a profile drawn to true earth radius is shown in Fig. 2. The survey of the path was made by plotting the profile from geodetic maps which were available from Shreveport to the Louisiana-Texas border. The balance of the path was plotted from hand altimeter readings taken along the path survey line from the Louisiana-Texas border to Latex. All high points along the path were later rechecked by sensitive altimeter for accuracy. Tree covering along the path ranged from sparse to moderately heavy, principally deciduous with a few pine. Tree heights were determined on the high points of the path by measuring the height of the tallest tree in the group under consideration. No sizable streams or bodies of water were under the path. From Fig. 2 it should be noted that the path is roughly grazing at three points.

The test was conducted from May 20, 1953, through September 30, 1954. The weather for the period was generally normal except for precipitation. The month of May 1954 showed the only significant excess, with only 2 other months showing any excess. The balance of the period was so dry that the total deficiency was 17.58 inches. An average of the highest daily humidity readings which normally occur at or near sunrise showed an increase of about 3 per cent above normal. The lowest daily humidity readings usually occurring in the middle or late afternoons showed an average of about 6 per cent below normal. Temperatures for the period averaged nearly normal, with the days slightly warmer and the nights slightly cooler. Wind velocities were only slightly greater than usual with no significant variation from normal.

Fig. 3 shows the antenna installation at the Shreveport terminal. The paraboloid on the right side of the tower crossmember was illuminated by the 1,965-mc horizontally polarized dipole antenna. The paraboloid on the left was illuminated by the 955.5-mc vertically polarized dipole. The paraboloid to the left of the tower near ground level was illuminated by the 6,730-mc horizontally polarized horn-type antenna. Energy

---

Paper 56-968, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted May 28, 1956; made available for printing August 1, 1956.

H. R. MATHWICH is with the Radio Corporation of America, Camden, N. J.; E. D. NUTTALL and A. M. RANDOLPH are with the United Gas Corporation, Shreveport, La.; and J. E. PITMAN is with the Philco Corporation, Philadelphia, Pa.

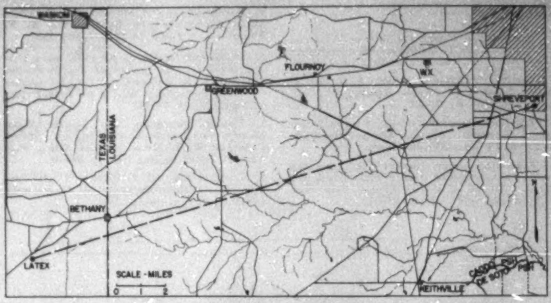


Fig. 1. Strip map of test path

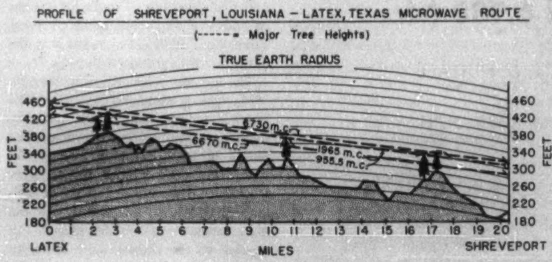


Fig. 2. Profile map of test path

was beamed from this parabola to the passive reflector at the top of the mast. The paraboloid near ground level in front of the tower and the passive reflector beneath the tower platform were part of the diversity tests not reported at this time. It is interesting to note that no significant loss differences were measured between horizontal and vertical polarization checks of the 955.5-, 1,965-, and 6,730-mc circuits. Antenna gains and circuit losses are a part of Table I. Fig. 4 is a view of the radio-equipment installation at the Shreveport terminal, showing, from left to right, the 1,965- 6,730-, and 955.5-mc radio-frequency equipment. The Latex antenna and equipment installations were similar to those at Shreveport.

The radio equipment used in the test was of standard manufacture. Since modulation was not used and bandwidth was not a consideration, the RCA equipment was adjusted where possible to increase its recording range. Modifications made to the Philco equipment did not increase the recording range over that obtainable from production equipment. Suitable arrangements were made for attaching signal recording instruments to the receiver for each test frequency.

Table I. Propagation Tests

	Circuit Frequency, Mc		
	6,730	1,965	955.5
Equipment	Philco *CLR-6*	RCA *CW-20A*	RCA *CW-5A*
Transmitter location	Latex	Latex	Latex
Transmitter power output, watts (db referred to 1 watt)	1 (0)	3 (+5)	2 (+3)
Receiver location	Shreveport	Shreveport	Shreveport
Paraboloidal antenna size, feet	6	6	4
Passive reflector size	6 by 8 feet		
Nominal antenna gains, two terminals, db	79.0	59.4	39.6
System loss for free-space transmission, db	66.3	77.0	90.8
Over-all median loss, db	80.6	92.0	105.1
Assumed obstruction loss, db	14.3	15.0	14.3

Technical Methods

The received signal level of each circuit was recorded continuously for the test period. The following information for each 15-minute period was transcribed manually from each circuit's daily chart to its daily data sheet:

1. Average loss value in decibels (db).
2. Minimum loss value in db.
3. Maximum loss value in db.
4. Number of times loss was greater than 20 db above reference value.
5. Duration of time loss was greater than 20 db above reference value, in minutes.
6. Number of times loss was greater than 30 db above reference value.
7. Duration of time loss was greater than 30 db above reference value, in minutes.
8. Number of times loss was greater than 40 db above reference value.
9. Duration of time loss was greater than 40 db above reference value, in minutes.
10. Equipment condition (one of seven symbols to denote equipment normal, under maintenance, etc.).

From the data sheets the information was punched on International Business Machines Corporation (IBM) cards, one card for each 15-minute period for each circuit. IBM machines were used for sorting and tabulating the data, but the mathematical computations and curve drawing were done by hand.

The receivers and transmitters for each circuit were calibrated as nearly as possible on a weekly basis and also before and after maintenance. All calibrations were made in db referred to one watt. Accuracy of the calibrating equipment was ±2 db. To provide a receiver calibration for a given period, a calibration curve was plotted from values averaged from the calibrations made at the beginning and end of that period. This curve was corrected to the standard transmitter output of 1 watt or 0 db by using the average transmitter power for the period as a correction factor. The corrected curve was then used to prepare a chart-reading card for transcribing the data from the charts to the data sheets.

The speed of the Esterline-Angus recording paper was chosen as 3 inches per hour as a practical maximum. For any particular chart speed, a certain minimum fade duration can be defined due to the finite width of the pen's trace. For this chart speed, a line in a downward direction followed by a return in an upward direction 1 minute later are barely separable from each other. For this reason, and because fades of shorter than 1-minute duration were expected to con-

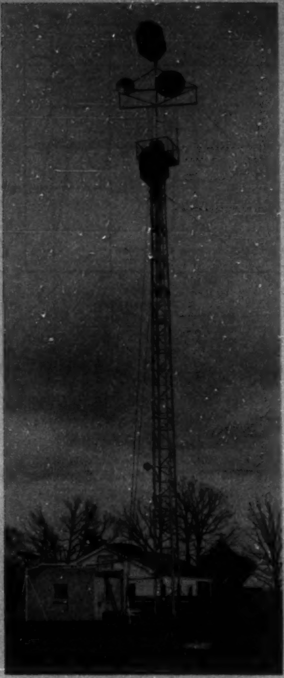


Fig. 3. Shreveport antenna installation

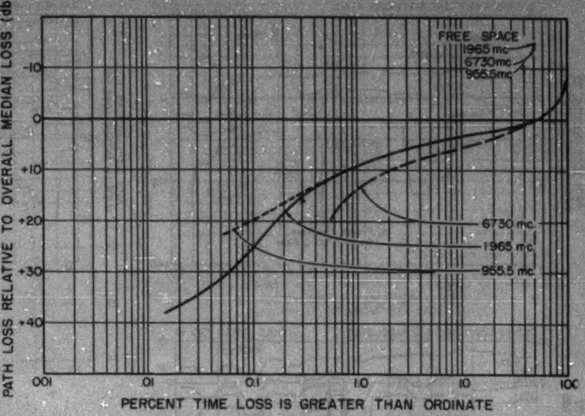


Fig. 5. 955.5-mc, 1,965-mc, and 6,730-mc loss distribution, May 20, 1953, through September 30, 1954

tribute very little to the total outage time, all fades of 1 minute or less were tabulated as 1-minute fades. It was proved later that this assumption had only a very small effect upon the over-all statistics. For example, this assumption would be expected to have the greatest effect on the statistics of the 6,730-mc circuit since these have the shortest median fade lengths [see Figs. 8(A), (B), and (C)]. From Fig. 9(A) it can be seen that only 15 per cent of the total outage time was contributed by fades of 1 minute or less duration at the +21-db level. Thus, if this total time were in error by a factor of two it would only change the total outage time for the +21-db point on the over-all distribution curve by 7.5 per cent. Consulting Fig. 5, this equivalent inaccuracy of level would make these data pessimistic by about 1 db. Similar analyses have proved none of the rest of the data to be in question by more than this 1 db.

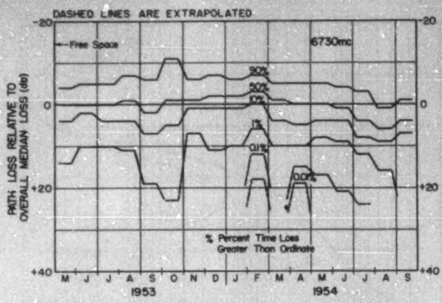
At the start of the test it was felt that the major item of importance was the time the loss increased to fairly high values. For this reason actual integration of the outage time was set up only for losses of 20, 30, and 40 db with respect to the selected reference losses. As the test progressed, however, it was felt that following the variations of the median value and other values near it also would be necessary. It then was reasoned that since the signal variations when the signal is near the median value are usually slow, it would be acceptable to use the 15-minute average signal. Thus, as an example, the outage time for the 4-db point used was the sum of the number of cards whose average loss was greater than 4 db times 15 minutes.

Definition of Terms

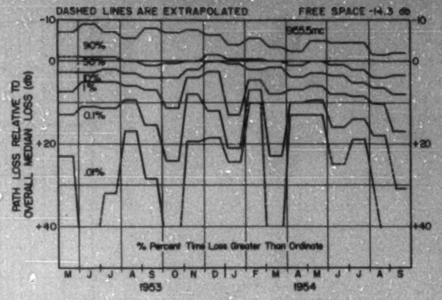
"Median loss" is that loss between transmitter output and receiver input such that 50 per cent of the time the observed loss was less, and 50 per cent of the time the loss was greater. In the case of "over-all median loss," this definition applies to data for the entire test.

"System loss for free-space transmission" is the loss that would be measured between transmitter output and receiver input if the path were operated in free space without obstructions between the antennas. This loss was computed by the equipment manufacturers who used transmission-line losses measured at the test site and specification gains of the antenna systems.

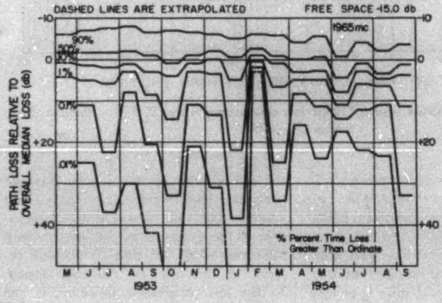
Fig. 4. Shreveport radio equipment installation



A—6,730 mc



C—955.5 mc



B—1,965 mc

Fig. 6. Seasonal probability May 20, 1953, through September 30, 1954

"Assumed obstruction loss" is the difference between over-all median loss and system loss for free-space transmission and represents the loss nominally attributed to the obstructions in the path.

Data

The primary information desired from the test was the loss distributions of the three frequencies involved operated over the same path. In setting up the test, consideration was given to providing other data such as seasonal and diurnal variations, fade-duration and faded-time distributions, rate-of-fading, and weather correlations. Since the difficulty of correlating fading and weather effects was well known, no specific and detailed effort was made in the primary phase of the test to make such correlations. A Weather Bureau station located on the path provided a source of weather data already punched on IBM cards on an hourly basis (Weather Bureau *WBAN-1* cards). With the test data also on IBM cards, it is expected that this aspect of the test will be completed as time permits.

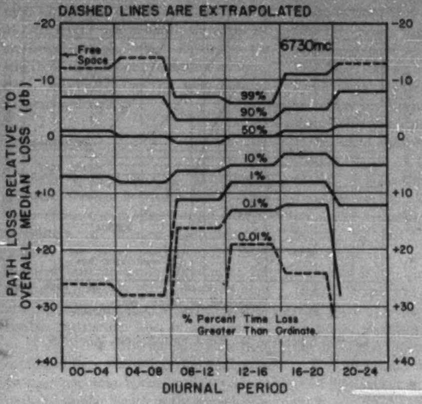
Loss Distribution Curves

The curve shown in Fig. 5 is perhaps of the greatest significance. In this curve the loss distributions for the three frequencies are shown. In general, it shows that as the frequency increases, the amount of time the loss is greater than a selected value increases. It must be cautioned, however, that data from this figure should not be used alone to judge the relative quality of the different frequencies as regards their use in microwave relay systems or their use in different geographical areas. Other items which must be considered include antenna gain, transmission-line or wave-guide loss, radio-frequency output availability, and attainable receiver noise figures.

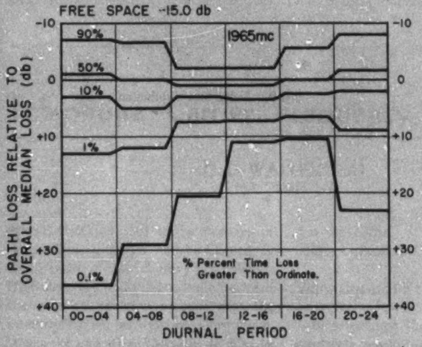
Seasonal Probability Curves

In the early phases of the test it was hoped that the seasonal variations noted, Figs. 6(A), (B), and (C), could be correlated with seasonal changes in K, the effective earth's radius, since K does vary on a rather periodic basis. Computations of the expected variation in signal were made assuming grazing smooth sphere diffraction. These computations predicted that the greatest losses would be present in February and the least in August. The calculated difference between the February and August median losses was 1.0 db for the 955.5-mc circuit, 1.4 db for the 1,965-mc circuit, and 2.3 db for the 6,730-mc circuit. It can be seen from inspecting the data that seasonal effects of a random nature are present to such a degree that even if the calculated effect were present, it would take many years of data to enable its extraction from the random data. For example, both the summers of 1953 and 1954 were covered by this test, but the mean signals of the summers were quite different, as shown in the median losses.

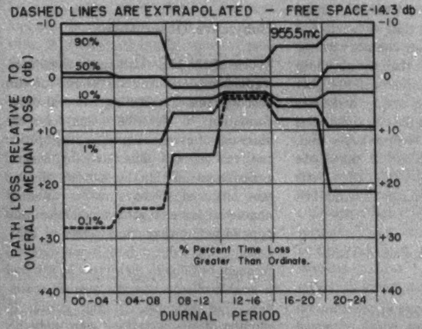
It can be noted, however, that the variation in median loss from one month to the next increases as the frequency increases. This accords with theoretical expectations that a given change in path obstruction height will cause an increasing change in the value of the received signal as the frequency becomes higher.



A—6,730 mc

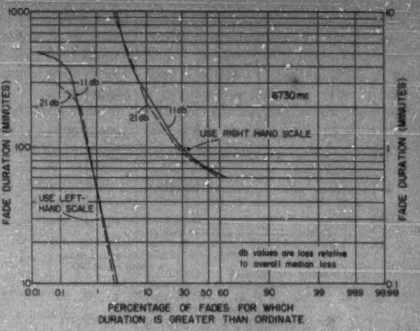


B—1,965 mc

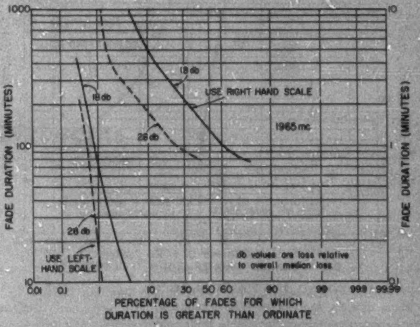


C—955.5 mc

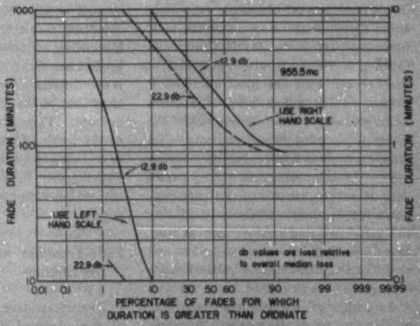
Fig. 7. Diurnal probability May 20, 1953, through September 30, 1954



A—6,730 mc, Dec. 1, 1953, through Sept. 30, 1954

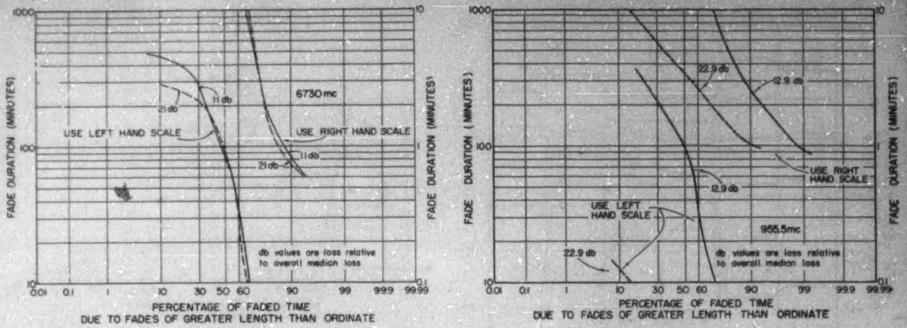


B—1,965 mc, May 20, 1953, through Sept. 30, 1954



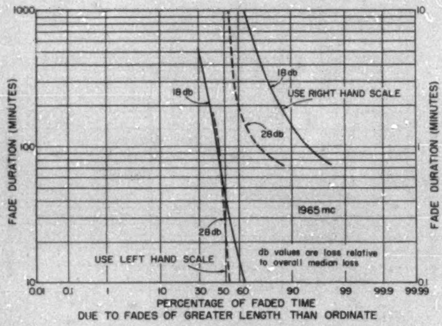
C—955.5 mc, May 20, 1953, through Sept. 30, 1954

Fig. 8. Fade-duration distribution



A—6,730 mc; Dec. 1, 1953, through Sept. 30, 1954

C—955.5 mc; May 20, 1953, through Sept. 30, 1954



B—1,965 mc; May 20, 1953, through Sept. 30, 1954

Fig. 9. Faded-time distribution

Diurnal Probability Curves

In Fig. 7, loss distributions collected into 4-hour diurnal blocks are presented. From these it can be seen that the most favorable part of the day (for minimum faded time) is in the afternoon; also, it can be seen that the most unfavorable time is in the 4-hour interval between midnight and 4 a.m. Presumably the afternoon is most favorable because, on an average day, the sun has been warming the earth for some time. This causes vertical turbulence which does not support stratification. Conversely, during the early morning hours winds and atmospheric turbulence tend to be low and stratification is much more liable to cause fading.

An additional observation is that the median loss on all circuits was found to be lowest near midnight. The probable explanation of this is as follows: On a typical night the evening starts with the ground relatively warm from the sun's radiation. As night falls, the ground radiates its heat energy into space, becoming cooler; however, the higher atmosphere does not do this. Therefore, a layer of air in immediate contact with the ground becomes cooler than the air above it and thus has a dielectric constant higher than that of the air above by a definite amount. This causes the effective radius of the earth (K) to become larger and, in a grazing path such as this, causes the signal levels to rise. This effect was noticed on many evenings to a marked degree.

Fade-Duration and Faded-Time Distribution

Two informative types of presentation of the observed fades which occurred during the test are the fade-duration distribution curves, Fig. 8, and the faded-time distribution curves, Fig. 9. It should be pointed out that in Figs. 8(A) and 9(A) the period of time covered is not the entire test period.

Subjective Observations

One of the important subjective observations made during the course of this test was that it takes a great deal of information to describe accurately the statistics of even one specific propagation path. This was illustrated by several occurrences. First, the summer of 1953 was observed to have rather different characteristics as regards median loss and type of fading than the summer of 1954. Second, the substandard fades were infrequent but due to their great length had a sizable effect upon the over-all statistics. For example, in the entire 16 months of the test there were only six fades of 30 minutes or longer duration for the 1,965-mc channel at the −18-db level. These fades, however, contributed 52 per cent of the total faded time at that level. Six is hardly an adequate statistical sample, which illustrates the length of time needed to obtain accurate data, especially on the long-duration substandard fades.

Conclusions

On the basis of the data taken during this test, the loss is greater than a given value relative to median loss for a larger percentage of time as frequency increases. In general, fades become more frequent but shorter in duration as the frequency increases. It can be seen that grazing paths of the type tested suffer a sizable obstruction loss, in this case 14 to 15 db. The 1,965-mc data have been compared for a 4-month period with other propagation data taken on 2,000 mc on paths of similar length in the same locale, but with adequate clearance to insure their median signals being near the free-space value. This comparison revealed that the grazing path was 30 db or more below free-space value three times longer than the clear path, and 40 db or more nearly five times longer. Caution should be used in considering the actual value of the reliability advantage. A comparison of the circuit outage time for the two hops showed the time advantage on the clear path to be approximately 3 hours for the 4-month period. This advantage was purchased for nearly three times the antenna cost. Land required for guying would cost about four times as much and the tower itself twice as much. A complete economic evaluation of the cost of additional propagation reliability is required to ascertain its actual value for any particular hop or system.

The importance of good propagation reliability for systems using telemetering, remote control, and supervisory functions, is apparent from an examination of the diurnal curves which show that periods of maximum loss occur during the periods of greatest need for control functions. It should be emphasized again that, in addition to the important factor of propagation reliability, other factors such as antenna gain, transmission-line loss, and economic aspects must be considered in the design of a system.

References

1. Propagation of VHF and SHF Waves Beyond the Horizon, K. Bullington. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 38, Oct. 1950, pp. 1221–22.

2. Selective Fading of Microwaves, A. B. Crawford, W. F. Jakes. *Bell System Technical Journal*, New York, N. Y., vol. 31, Jan. 1952, pp. 68–90.

3. Results of Microwave Propagation Tests on a 40 Mile Overland Path, A. L. Durkee. *Proceedings*, Institute of Radio Engineers, vol. 36, Feb. 1948, pp. 197–205.

4. A Microwave Propagation Test, J. Z. Millar, W. F. Byam. *Ibid.*, vol. 38, June 1950, p. 619.

5. Microwave Propagation Experiments, L. E. Thompson. *Ibid.*, vol. 36, May 1948, pp. 671–76.

Telephone Carrier Frequency Networks

G. D. WALLENSTEIN
NONMEMBER AIEE

THE GROWTH of carrier systems in the United States may be likened to the growth of the country's highways. Highways originally were built between the outskirts of cities to cope with the intercity traffic; so were carrier systems. When he traveled through a city, a motorist had to leave the highway and become submerged in local traffic until he could make connection to the next highway heading out of the city. Likewise, if carrier channels were to pass through a city, they had to become local voice drops first before they could be reinserted into another carrier system and passed on to the next city.

Highway engineers now not only build superhighways between cities, but also provide by-passes and through-connections by way of interchanges. The same can now be done with carrier systems, with interconnections and dropouts to suit the traffic. It is done with 45-class carrier networks.

Voice Channel Networks

In telephony, the ability to establish electrical connection between any two telephone instruments is the basic commodity-of-value to the subscriber. From local networks for a privileged few, the industry has grown to the point where global interconnection is taken for granted. In 1954, about 96 per cent of the world's registered telephones were potentially able to be connected one to another.

In general, such connections between any two of 90 million stations are random in character, unscheduled, and of temporary duration. This means that the connection has to be flexible and anonymous and that common standards must be employed in all participating telephone systems. Three basic features are required to accomplish these objectives.

1. Any two telephones must be able to form a working bridge for talker and listener, regardless of the instrument's origin and of the type of speech employed.

2. Telephone companies must provide some means of subscriber identification and of calling the selected party.

3. Local, regional, and national networks must have points of connection for the interchange of electric energy representing the intelligence to be transmitted.

During the first 40-odd years of telephony, telephone transmission relied on the use of direct voice frequencies as contained in speech. It confined itself to translating the acoustical energy into equivalent electrical phenomena and vice versa. Networks permitting connection between subscribers were required to handle and recognize these voice frequencies. Difficulties in maintaining standards of intelligibility over long and multisection connections, however, led to the need for definition of transmission requirements. Thus it is now possible to refer to the voice channel as a unit of information exchange upon which all telephone connection schemes are based.

It is noteworthy that discussion is still active regarding the minimum requirements for the voice channel although an international committee, Comité Consultatif International Telephonique (CCIF), founded in 1923, has deliberated on such questions for about 30 years. Nevertheless, within some margins of latitude on actual circuit quality, it is now standard practice to request and deliver voice channels fit for interconnection between a great number of separately controlled telephone companies. So much has the voice channel become an accepted

---

Paper 56-783, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 26, 1956; made available for printing April 30, 1956.

G. D. Wallenstein is with the Lenkurt Electric Company, San Carlos, Calif.

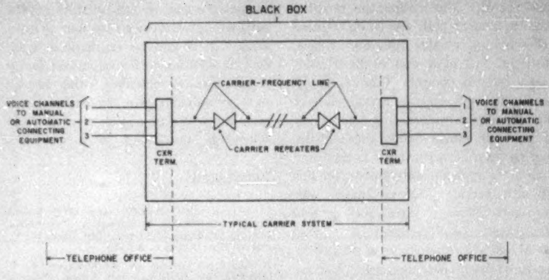


Fig. 1. The "black-box concept" applied to conventional carrier system applications

traffic unit that even the historically earlier telegraph transmission, now primarily conducted by modulation of alternating frequencies, is handled in channel increments which will conveniently fit into the capabilities of one voice channel at a time.

With basic agreement on the characteristics of the voice channel, and equipped with tools which make transmission practical over any desired distance (amplifiers, properly transposed open-wire lines, loaded cables, overseas radio), telephony came to a realization of unlimited interconnection in the late 1920's. The carrier frequencies then entered the picture. Today, in the United States alone, there are about 37 million circuit-miles of carrier telephone circuits, and the growth in this field is such that a saturation point is nowhere in sight.

The voice channel standard alone is no longer adequate for interconnection as wires and radio waves carry a multitude of speech messages encoded in arbitrary frequencies. How carrier frequencies can be standardized and integrated into interconnected network is the subject of this paper.

The Carrier-Frequency Telephone Plant

Carrier systems made their first appearance in the 1920's and have grown in use until the maxim that, "carrier is king," coined in 1949, is nearly literal truth.[1] In Europe, the CCIF has labored over problems of standardizing carrier system characteristics since the end of World War II, with participation from all parts of the world. Yet, even today, new carrier systems are being developed for specific telephone companies or for peculiar transmission conditions with little thought for network connections. It will be necessary to review the role of carrier in the telephone plant before an appraisal of network application can be made.

The entrance of carrier into telephony may be likened to other technological advances in communications. Just as the first engine on an old-fashioned coach produced literally a "horseless" carriage, so did carrier at first mean no more than "frequency-translated voice channels." With this concept, the desired end-result is still the same voice channel which, during its course of transmission, is subjected to arbitrary changes to allow a better utilization of the transmitting medium. The familiar "black-box" concept applies to everything contained between the two end-terminal positions of the voice channel. This principle is illustrated in Fig. 1.

A carrier system so applied provides for a simultaneous number of voice channels to connect two specific locations. With some types of carrier systems, each channel may be carried to a different location along the same route, as shown in Fig. 2. In either case, the job is completed at the voice channel terminal point. If connection requires, such a voice channel terminal may be switched or patched to another similar terminal, carrier-derived or physical-type.

With connection requirements applicable at the voice channel terminals only, there is no restriction on the particular *modus operandi* of the carrier system. This freedom has led to a multitude of approaches to carrier system operation. There exist today 4-wire and 2-wire carrier systems, single-side-band and some double-side-band types, systems with 3-, 4-, 5-, 6-, 8-, and 12-kc spacing between channels, systems which change direction of transmission in the same spectrum where others transmit only in one direction, etc. Lack of agreement on performance requirements for voice channels has further infiltrated carrier system design by permitting a wide range of quality both with respect to reproduction of frequencies and freedom from unwanted interference.

Fortunately there has been a growing tendency for voluntary regulation in the United States and Canada. Spearheaded by the American Association of Railroads and aided by the Bell System's very large interest in uniformity, there have been on record for some years now recommended frequency assignments for carrier systems.[2] Similar work is being done in Europe by the CCIF.[3] Such regulation has made it possible for many different carrier systems to exist on adjacent wires of the same pole line or cable. The requirement that carrier systems connect to each other at other than voice frequencies has only been applied to so-called backbone or international circuits.

Telephone companies are now adding toll channel facilities in increments of 4, 12, or more channels at one time. The

Vol. 7
Pt. 1
1956

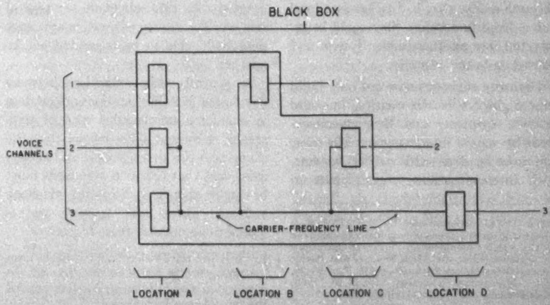


Fig. 2. A typical 3-channel stackable carrier system considered as a single black box

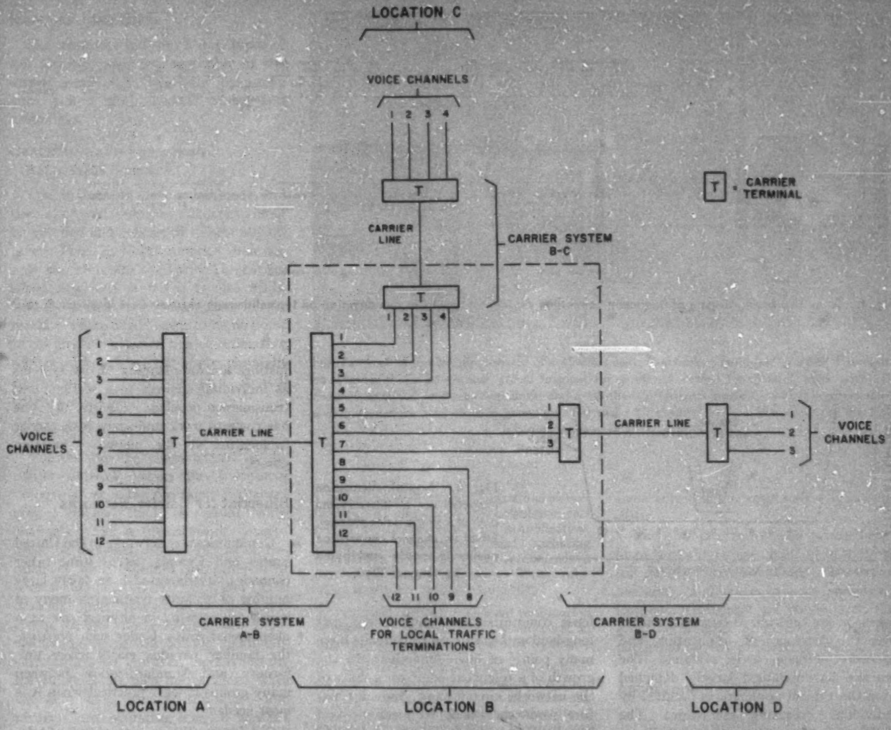

LOCATION C
VOICE CHANNELS
1 2 3 4
T
CARRIER LINE
CARRIER SYSTEM B-C
T
1 2 3 4
T = CARRIER TERMINAL
VOICE CHANNELS
1 2 3 4 5 6 7 8 9 10 11 12
T
CARRIER LINE
T
1 2 3 4 5 6 7 8 9 10 11 12
1 2 3
T
CARRIER LINE
T
1 2 3
VOICE CHANNELS
12 11 10 9 8
CARRIER SYSTEM A-B
VOICE CHANNELS FOR LOCAL TRAFFIC TERMINATIONS
CARRIER SYSTEM B-D
LOCATION A
LOCATION B
LOCATION D

Fig. 3. Typical connections between individual voice channels of separate carrier systems

toll network is becoming such a mesh of direct, alternate routes and cross-connections that carrier systems are no longer laid out to provide end-traffic between two final points. At numerous points of connection, some of the channels combined in a given carrier system may branch off and others may terminate locally, while the remainder may go on to a more distant point served by another carrier system in a similar fashion. Such situations can be handled, of course, by provision of separate carrier systems for each section between drop-off or branching points, and connection of the voice-channels back-to-back, one at a time, as required (see Fig. 3). This has been the most common practice outside the backbone toll routes; however, it has severe technical and economical disadvantages.

From a technical standpoint, every voice-frequency connection between carrier systems repeats the carrier-modulation procedure and thus adds to the number of times the frequency accuracy and range of the derived voice channels may be affected, as illustrated in Figs. 4 and 5. Economically, the major cost factor in carrier systems is the equipment associated with the channelizing process (modulation and filtering), and the attendant expense in office mounting space, power, and maintenance. If carrier systems can be so constructed that the channelizing elements may be omitted, their direct through-connections can be made for bands of carrier frequencies which carry a number of voice channels and more than 50 per cent of the cost can usually be saved. The connection diagram of Fig. 4 would then appear greatly simplified as shown in Fig. 6.

For carrier systems to become suitable for network layout without resorting to conventional voice frequencies, three requirements must be met:

1. The channel end and the carrier-frequency transmission equipment must be separable.
2. Specific increments of channels must be recognized as group units at standardized frequencies.
3. All commonly encountered transmission media must qualify for system application.

Such requirements amount to a master plan for telephone transmission networks which is comparable to other fundamentals such as the common railroad gauge which permits the same car to travel from one end of North America to the other. It is not surprising, therefore, that until recently this concept had only been adopted in two limited cases: in the United States, by the American Telephone and Telegraph Company for backbone circuits served by the *J/K/L* carrier systems; and in Europe, on selected international toll circuits under CCIF recommendations.[4]

In the United States, advantage was

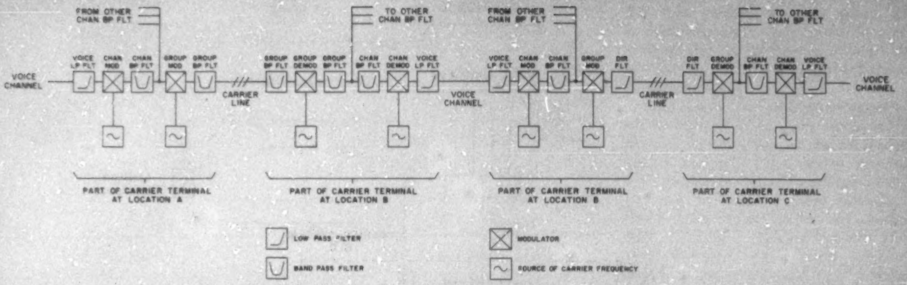

FROM OTHER CHAN BP FLT
TO OTHER CHAN BP FLT
FROM OTHER CHAN BP FLT
TO OTHER CHAN BP FLT
VOICE CHANNEL
VOICE LP FLT
CHAN MOD
CHAN BP FLT
GROUP MOD
GROUP BP FLT
CARRIER LINE
GROUP BP FLT
GROUP DEMOD
GROUP BP FLT
CHAN BP FLT
CHAN DEMOD
VOICE LP FLT
VOICE CHANNEL
VOICE LP FLT
CHAN MOD
CHAN BP FLT
GROUP MOD
DIR FLT
CARRIER LINE
DIR FLT
GROUP DEMOD
CHAN BP FLT
CHAN DEMOD
VOICE LP FLT
VOICE CHANNEL
PART OF CARRIER TERMINAL AT LOCATION A
PART OF CARRIER TERMINAL AT LOCATION B
PART OF CARRIER TERMINAL AT LOCATION B
PART OF CARRIER TERMINAL AT LOCATION C
LOW PASS FILTER
BAND PASS FILTER
MODULATOR
SOURCE OF CARRIER FREQUENCY

Fig. 4. Simplified block diagram of frequency-controlling devices employed in one direction on typical through channels from locations A to B to C of Fig. 3

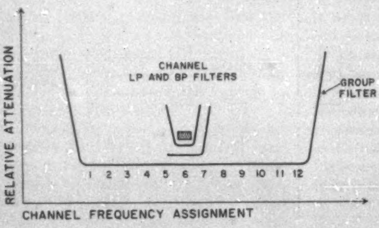


Fig. 5. Typical attenuation characteristics of filters in the path of a single voice channel. Effect on channel transmission quality is mostly contributed by channel filters

taken of one centralized common control over the development, application, and operation of these carrier systems. The practice has remained largely detached from the typical applications handled by individual operating companies. The CCIF recommendations have the earmarks of restrictive legislation carefully limited to applications of international scope where such procedure can be justified. Though both cases thus have failed to include the regional and general short-haul users of carrier systems, the development of a nation-wide toll dialing system and the network coverage required by the SAGE comprehensive defense communications have shown that long-haul and short-haul toll circuits have many points of interconnection. In the growth of a telephone company's share of the network, circuits may gradually mature from short-haul to medium- and long-haul extensions. Carrier systems of the greatest concentration of voice channels may often develop at the short-haul toll end links.

Short-haul carrier systems are no longer of the low-channel capacity type, a characteristic which had led to the working assumption that such systems are less critical and require less care in planning and operation. There is less and less justification for planning carrier systems as individual answers to a sectionalized transmission problem. Clearly, the time has come to work uniformly with carrier frequencies at the network junction points.

Blueprint for Carrier Networks

Communication networks in the United States and Canada, as in some other countries, are operated by a fairly large number of separate companies, many of which are growing in physical size at a phenomenal rate. Under such pressure, the familiar paradox exists where uniformity and standardization between many groups is least practical when it is most needed.

Considering the present growth of telephone companies, it seems natural for a carrier system manufacturer to attempt to satisfy the needs of all these users by thinking in terms of their network requirements and by searching for a common solution acceptable to all. Thus, a master plan was developed (since introduced as the "*45* Concept") based upon the following considerations.

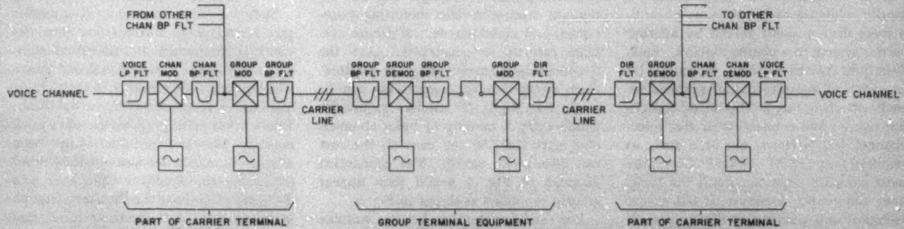

FROM OTHER CHAN BP FLT
TO OTHER CHAN BP FLT
VOICE CHANNEL
VOICE LP FLT
CHAN MOD
CHAN BP FLT
GROUP MOD
GROUP BP FLT
CARRIER LINE
GROUP BP FLT
GROUP DEMOD
GROUP BP FLT
GROUP MOD
DIR FLT
CARRIER LINE
DIR FLT
GROUP DEMOD
CHAN BP FLT
CHAN DEMOD
VOICE LP FLT
VOICE CHANNEL
PART OF CARRIER TERMINAL AT LOCATION A
GROUP TERMINAL EQUIPMENT AT LOCATION B
PART OF CARRIER TERMINAL AT LOCATION C

Fig. 6. Simplified block diagram of same network as in Fig. 4, but with carrier-frequency interconnection at B. The net saving of this arrangement over that in Fig. 4 is one half of the channel equipment for each voice channel included in the connection

NETWORK FACILITIES

The telephone industry's four types of toll facilities must all be usable in the carrier network. These are open-wire lines, paired cables, coaxial cables, and radio links.

CARRIER-SYSTEM FREQUENCY-ALLOCATION SCHEMES

The investment in predominant and time-approved frequency allocations must be guarded and continued to best advantage. Thus, in North America there are now two standard open-wire line carrier allocations, one of which is also traditionally preferred in other parts of the world. There would seem to be no point to the creation of a new non-co-ordinating scheme although a fresh approach might conceivably be more convenient from a network standpoint.

CHANNEL GROUPS

The outstanding characteristic of the carrier network versus the traditional voice network is the provision of channel groups rather than individual voice channels. The entire network concept must be designed around blocks of frequencies which contain transmission of several channels. The $J/K/L$ and CCIF systems have settled on a 12-channel group located at 60–108 kc, and a 60-channel group at 312–552 kc. As in any standardization, however, there is nothing inevitable about such frequencies. With the field wide open at the short- and medium-haul operating level, it was felt that groups could be established which would permit the least cumbersome modulation techniques for line transmission in the most widely used carrier system allocations. In addition, a group with less than 12 channels seemed necessary, as evidenced by a large-scale use of 3- and 4-channel carrier systems.

The groups finally evolved were the following:

1. A 4-channel block of 8–24 kc, termed a "pregroup."
2. A 12-channel block of 40–88 kc, called a "basegroup" in conformance with CCIF terminology.
3. A 24-channel block at 40–140 kc, termed a "double basegroup."

Larger groups may be adopted as the need arises, with the established 60-channel "supergroup" of 312–552 kc the next logical increment.

EQUIPMENT BLOCK-BUILDING DESIGN

To get all the benefits of carrier network operation, the equipment must be arranged in separate blocks as follows:

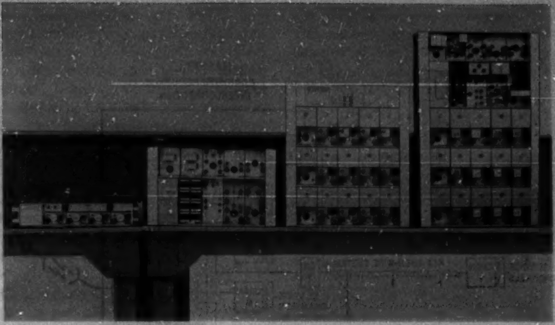


Fig. 7. Four of the building blocks of a 45-class carrier network. From left to right: Four-unit compandor shelf, 24-channel group terminal for multipair cable, 12-channel bank with 12 channel units and four pregroup units, and a 12-channel terminal assembly which is effectively a combination of a 12-channel group terminal and a 12-channel bank. The height of the 12-channel terminal is 31 1/2 inches

1. Equipment for a particular transmission facility with the effects of transmission corrected by the processes of amplification, regulation, equalization, and including modulation to provide group frequencies of one of the standard groups. This equipment is called a "group terminal."

2. Channel banks for the local termination of groups of voice channels, connecting in this case to the associated group terminals.

3. Optional equipment blocks as required for connecting group terminals together, and for voice channel drop requirements. In this class also are group increments of channel compandors, a device required only on certain transmission facilities for toll quality from a noise standpoint. Further included is equipment to connect the American Telephone and Telegraph and CCIF type of 60–108 kc transmission equipment where required. Typical major equipment blocks are illustrated in Fig. 7.

EQUIPMENT STANDARDIZATION

Less than optimum network economy would be realized if each of the equipment blocks used was unique and peculiar. It has been possible to subdivide the major building blocks into types of units which can be repetitively used. From a maintenance and circuit reassignment standpoint, the possibilities of the carrier network concept are perhaps equal to or greater than the immediate benefits at planning and installation time.

In a telephone toll office where open-wire line, cable carrier, and perhaps some radio systems terminate, it has been hitherto typical to service each carrier terminal equipment with completely different spare parts and instructions. However, the new channel banks employ the same units on the terminals for any facility.

This standardization also has the benefit of providing voice channels of uniform toll quality regardless of the transmission medium. Differences in short-haul versus long-haul treatment are confined to the group terminal equipment tailored to the particular facility's operating scheme.

Applications

Characteristic applications for carrier networks occur on toll telephone trunks between cities partly connected by open-wire and partly by conventional voice-frequency balanced cable. Such an application was described in an earlier paper.[5]

One particularly fertile field of application is the rapidly growing use of radio for toll routes. As radio systems operating at microwave frequencies tend to permit simultaneous transmission for several hundred voice channels, it is natural to utilize branching and local dropping of carrier frequency channel groups on such systems. Moreover, many radio installations are located away from the city for space propagation reasons, with long wire-line or cable connections to the toll office. With carrier network interconnection equipment, the city-to-radio-station transmission can be handled on a co-ordinating, standard basis with no channel terminal equipment required at the radio site.

An actual example of a typical complex network application is illustrated in Fig. 8. This application involves the establishment of a new radio toll route and of new

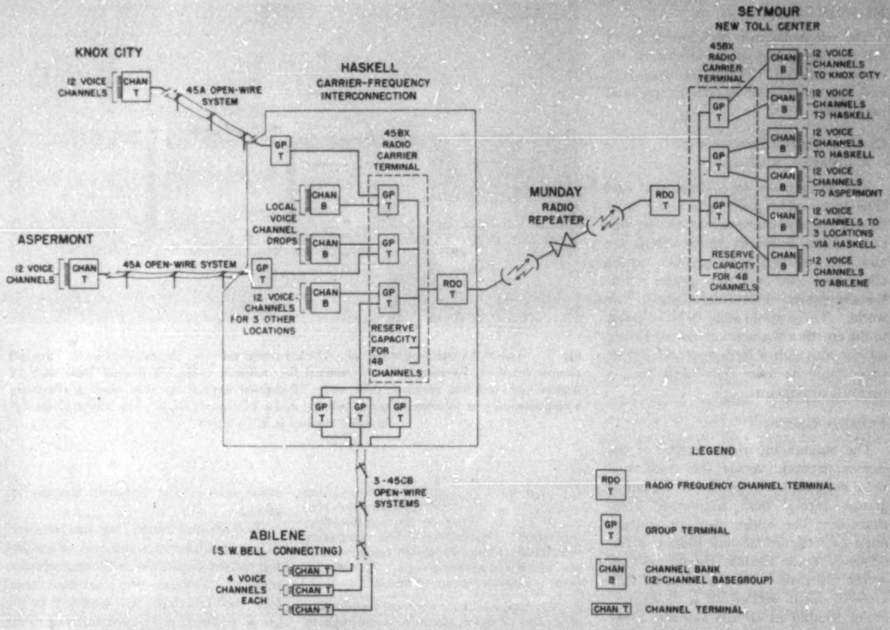

SEYMOUR
NEW TOLL CENTER
45BX RADIO CARRIER TERMINAL
GP T
CHAN B
12 VOICE CHANNELS TO KNOX CITY
12 VOICE CHANNELS TO HASKELL
12 VOICE CHANNELS TO HASKELL
12 VOICE CHANNELS TO ASPERMONT
12 VOICE CHANNELS TO 3 LOCATIONS VIA HASKELL
12 VOICE CHANNELS TO ABILENE
RESERVE CAPACITY FOR 48 CHANNELS
RDO T
KNOX CITY
12 VOICE CHANNELS
CHAN T
45A OPEN-WIRE SYSTEM
HASKELL
CARRIER-FREQUENCY INTERCONNECTION
45BX RADIO CARRIER TERMINAL
LOCAL VOICE CHANNEL DROPS
12 VOICE CHANNELS FOR 3 OTHER LOCATIONS
RESERVE CAPACITY FOR 48 CHANNELS
MUNDAY
RADIO REPEATER
ASPERMONT
12 VOICE CHANNELS
45A OPEN-WIRE SYSTEM
3-45CB OPEN-WIRE SYSTEMS
ABILENE
(S.W. BELL CONNECTING)
4 VOICE CHANNELS EACH
LEGEND
RDO T RADIO FREQUENCY CHANNEL TERMINAL
GP T GROUP TERMINAL
CHAN B CHANNEL BANK (12-CHANNEL BASEGROUP)
CHAN T CHANNEL TERMINAL

Fig. 8. Proposed toll network plan for the Seymour-Haskell, Tex., area served by the General Telephone Company of the Southwest

open-wire carrier systems in the territory of the General Telephone Company of the Southwest. At present, the exact layout is tentative and still subject to change.

The telephone company has the problem of retiring a presently existing manual toll center as a part of a dial conversion program. Voice channel concentration between the new combined toll center at Seymour, Tex., and the distribution point at Haskell, Tex., is large enough to favor the economy of a microwave radio system.

From Haskell, which will become an unattended office, channel groups from the radio terminal will be directed to their respective assignments on a number of radiating open-wire leads. The characteristics of these leads are such that they will involve the use of *45A* carrier systems (*J*-co-ordinating), *45C* carrier systems (*O*-co-ordinating), and voice-frequency circuits. A total of seven different locations, including Haskell, will be connected with the Seymour office. Four of them will be served directly by the *45*-class carrier network with carrier-frequency interconnections at Haskell. The remaining three towns will continue to be served by voice-frequency interconnection at Haskell as there are suitable existing facilities.

Many networks such as the one shown in Fig. 8 are now in the planning and installation stage, and will unquestionably increase in numbers as more of the various equipment building blocks become available. There will be great need for many 4- to 24-channel direct, high-usage groups in the continental toll-dialing network.

Considering all possible applications of the carrier network concept, however, it is apparent that some logical limit will be encountered for the number of times the same channel group can be expected to enter and withdraw from connecting points. As a generalization, the frequency of interconnection of one and the same group would increase inversely with system length. Short-haul trunk groups may be expected to run through as many as six tandem carrier connections. With respect to over-all group-transmission quality, the group transmitted through many interconnections will perform about equally as well as a single long-haul system with many repeaters.

There are and will continue to be applications for hitherto conventional-type carrier systems, where the voice-channel requirement is very low or where interconnection with other channel groups is not usually intended unless in steps of one channel at a time. Another application of conventional-type carrier is in the exchange and subscriber telephone plant where the use of carrier is still in its infancy. At the present time it is in toll or similar nontelephone long-distance traffic that network carrier will find most use.

In the step-by-step development of a program such as outlined here, judgment must be constantly exercised to arbitrate between the pure concept and the pressure for prompt field application. Idealistically, it might be proper to develop the complete line of carrier network equipment with all offshoots for standard applications in one gigantic effort, if the industry would stand still long enough to settle on the range of applications known at the outset of such a program. Such an approach is not realistic, however, and might very well impede further advances of the art. By contrast, the concept of carrier networks may be considered timeless within reason if flexibility and compromise are used in its implementation.

A step-by-step approach has permitted piecemeal releases of systems and equipment building blocks, creating a skeleton of potential networks. Some equipment items have appeared which are combinations of building blocks, serving the same concept by elimination of plug-in units rather than by original reduction in effort to the smallest possible block increment.[6] In this third year of established field use, equipment for some of the open-wire applications is still not available, as preference had to be given to group terminal equipment for radio and cable.

From a technological standpoint, the experience of the past year has already shown that such a concept is fully compatible with the idea of progress in the communications art. For example, when the all-carrier network equipment was conceived, transistors were known only as interesting specimens for experimentation; hence, the first equipment made available was designed around vacuum tubes. Now, as additional equipment items are under development, transistors are used in preference to vacuum tubes. Future techniques of mass production and equipment simplification need not conflict with the standard group frequencies and interconnection techniques. These standards are of such a classic simplicity that their usefulness is assured for a long time to come.

Conclusions

The universal transmission of speech- and voice-channel-type pulses requires networks in which carrier frequencies rather than voice frequencies must be considered. A master plan for carrier-frequency operation in all types of communication networks provides the desired standards of connection coupled with greatest practical economy in equipment. This network concept makes carrier at home on all practical facilities and leaves the operating company free to plan to traffic requirements instead of to equipment limitations.

References

1. Carrier Is King, C. M. Mapes. *Bell Telephone Magazine*, New York, N. Y., vol. XXVII, no. 4, winter 1949/1950.

2. American Association of Railroads Manual. Communications Section. *Specification 2-J-46* adopted June 1946, revised Apr. 1953.

3. Green Book of the CCIF. Comité Consultatif International Telephonique, International Telecommunications Union, Geneva, Switzerland, vols. 3 and 4, 1956.

4. Programme General d'Interconnexion en Europe, et dans le Bassin Mediterraneen, 1952–1954. *Ibid.*, 1954.

5. Combined Operation of a 24-Channel Cable Carrier and 12-Channel Open-Wire Carrier Systems, G. W. Searle R. A. Schaefer. *AIEE Transactions*, vol. 74, pt. I, 1955 (Jan. 1956 section), pp. 770–77.

6. Application and Transmission Features of a New 12-Channel Open-Wire Carrier System, K. E. Appert, R. S. Caruthers, W. S. Chaskin. *Ibid.*, vol. 73, pt. I, Mar. 1954, pp. 18–27.

Lightning Protection on the Stevens Point—Wisconsin Rapids Intercity Telephone Cable

R. C. DOWLING
ASSOCIATE MEMBER AIEE

Synopsis: It has been a common assumption that buried telephone cables are practically free from the destructive effects of lightning, particularly intercity cables which have no intermediate terminals or connections to local subscriber installations along the route of the buried section. Such cables in some areas, however, have been found to be relatively vulnerable to damage from lightning strokes occurring along the route of cables. To avoid such damage various methods have been employed. One method has been to plow shield wires in the ground above the location of the cables. Whether or not such protection is provided is customarily based on balancing the cost of such measures against the expectancy of lightning strokes per year along the respective route, and the anticipated cost of repairs and interruptions entailed. This paper describes a situation where the anticipated troubles did not appear to warrant the cost of the initial installation of shield wires, but later experience indicated the need for some remedial measures. The measure adopted consisted of the installations of full-count carbon gap protection at various intervals along the cable route.

DURING 1946 an intercity telephone cable was placed between Stevens Point and Wisconsin Rapids. These small cities are centrally located in the State of Wisconsin. The cable followed the route of the Stevens Point–Minneapolis coaxial cable from Stevens Point for a distance of approximately 10 miles. Of this distance, approximately one mile is in conduit in the city of Stevens Point, and the remainder is aerial on the same poles with the coaxial cable. Beyond this aerial portion, approximately 9 miles are buried about 30 inches in the ground. Another length of approximately 1/2 mile is in conduit in the city of Wisconsin Rapids. Approximately 2 miles south from the junction of the aerial and buried cable, a terminal serves one phantom group of open wire, extending approximately 1/2 mile to the village of Rudolph. This terminal was equipped with a protector providing 700-volt gaps to sheath on the pairs in use. Fig. 1 shows the route of this cable and other pertinent data.

Cable Specifications

The cable is made of 19 quads of 19-gauge copper conductors, which are paper insulated and enclosed in a lead sheath. A double paper wrap between the conductors and sheath provides a dielectric strength of approximately 4,000 volts. The part of the cable plowed directly into the earth has a double serving of jute impregnated with a preservative compound over the lead sheath. The aerial part and that part in conduit have no protection over the lead. The logitudinal resistance of the lead sheath is 2.38 ohms per mile.

Preliminary Studies

In intercity cables, lightning damage normally occurs because the lightning current raises the potential of the sheath sufficiently above that of the conductors to cause insulation breakdown to some

---

Paper 56-971, recommended by the AIEE Wire Communications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted August 2, 1956; made available for printing August 2, 1956.

R. C. Dowling is with the Wisconsin Telephone Company, Madison, Wisc.

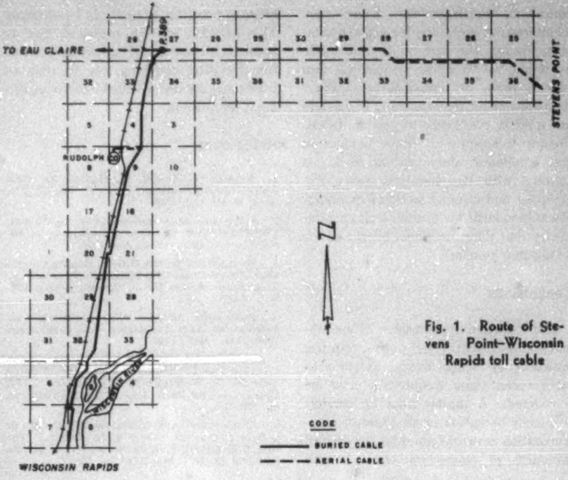

TO EAU CLAIRE
P. 389
STEVENS POINT
RUDOLPH
CO
WISCONSIN RIVER
WISCONSIN RAPIDS
N
CODE
BURIED CABLE
AERIAL CABLE

Fig. 1. Route of Stevens Point–Wisconsin Rapids toll cable

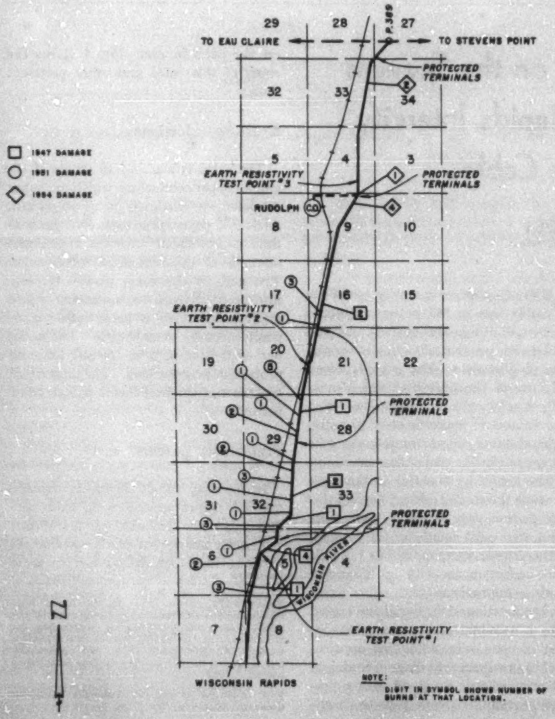

TO EAU CLAIRE
P. 389
TO STEVENS POINT
PROTECTED TERMINALS
EARTH RESISTIVITY TEST POINT #3
RUDOLPH
CO
PROTECTED TERMINALS
EARTH RESISTIVITY TEST POINT #2
PROTECTED TERMINALS
PROTECTED TERMINALS
WISCONSIN RIVER
EARTH RESISTIVITY TEST POINT #1
WISCONSIN RAPIDS
N
☐ 1947 DAMAGE
○ 1951 DAMAGE
◇ 1954 DAMAGE
NOTE:
DIGIT IN SYMBOL SHOWS NUMBER OF BURNS AT THAT LOCATION.

Fig. 2. Buried toll cable between pole 389 and Wisconsin Rapids

of the conductors near the sheath. These conductors are then raised in potential and may break down to other conductors at scattered locations wherever the insulation is the weakest. The usual mitigative procedure is to limit the sheath potential rise by shield wires or to increase the insulation between the sheath and conductors. These measures can best be applied in the design of new cable installations.

Before the installation of the cable, a study of the probable lightning hazards was made. Earth resistivity tests indicated the average resistivity to a depth of 40 feet to be 400-meter-ohms at point no. 1, 118-meter-ohms at point no. 2, and 54-meter-ohms at point no. 3. The locations of these test points are shown in Fig. 2. Note that each test point is approximately in the center of a 3-mile section.

With the use of the lightning storm incidence for the area, the longitudinal sheath resistance, and the above-earth resistivities in curves available at the time, it was estimated that the following should be anticipated: 1.2 cases of lightning damage per 100 miles of sheath per year in the 400-meter-ohm section, 0.5 cases per 100 miles per year in the 118-meter-ohm section, and negligible damage in the 54-meter-ohm section. This can also be expressed as one case of damage every 28 years in the first section and one case every 67 years in the second section. At that time, it was estimated that shield wires would reduce the expected damage rate to 0.77 cases per 100 miles per year in the first section and 0.32 cases per 100 miles per year in the second section. On the basis of these computations, the cost of shield wires could not be justified.

Although this paper is not primarily concerned with the aerial portion of the cable, it is of interest to note that the aerial portion of this cable and the coaxial cable are protected by pole protection wires and a buried shield wire.

Initial Experience

The buried portion of the cable was plowed into the ground during the spring. Before it was spliced, lightning damage occurred in three different pieces. Accurate information as to the extent and location of this damage was not recorded.

During 1947 scattered minor burns in the buried portion of this cable occurred at 14 locations. These points are indicated

Vol.
Pt. 1
1956

Fig. 3. First protected terminal north of Wisconsin Rapids looking north

on Fig. 2. Although some of this damage was definitely new, it was thought that part of it might have been caused by not clearing completely the previous trouble. Accordingly, there was no great concern over this damage and no mitigative measures were considered at that time. No major lightning damage to the buried portion of this cable was reported during 1948, 1949, or 1950.

1951 Damage

During a storm on July 16, 1951, the cable was damaged severely. Trouble occurred at 39 scattered locations. An analysis of the damage indicated the cause was three or four separate strokes of lightning. Fig. 2 shows the general location of this damage.

The most severe stroke occurred on a small tree located approximately 5 feet from the cable. This tree was shattered so completely that only a stump about 4 feet high remained. A large hole was blown in the ground to the depth of the cable. The cable was burned completely in two with the ends left sticking up clear of the earth. There was no visible evidence above ground of lightning strokes at any other locations.

The tree involved was 3 to 4 inches in diameter and showed signs of having been injured when small. It apparently had been just a crippled sapling when the cable was placed and was not considered worth cutting down. It must have grown rapidly, however, because the debris found indicated that it had been 10 to 15 feet tall. There were no other trees near.

A survey of these damages shows a group of four burns located within a mile north of the point where the 118-meter-ohm resistivity test was made. Two of the burns experienced in 1947 also had been in this locality.

Another group of five burns occurred very close together about 1/2 mile south of the location of the 118-meter-ohm resistivity test. This group, however, involved only open and crossed pairs, and the cable sheath was not involved in any of the burns.

Burns were scattered a distance of over a mile in the vicinity of the stroke to the tree described in the foregoing. Some of these were single burns and others were in groups. Many of the burns involved only conductors, but others involved the sheath and could have been caused by separate strokes.

A final group of burns occurred a little less than 1/2 mile south of the location of the 400-meter-ohm earth resistivity test. Although only five burns were listed at this location, several other troubles, which were not located exactly, were cleared by replacing a 25-foot section of cable.

Further Damage Expectancy Studies

In view of this severe damage, the need for special protective measures again was examined. The earth resistivity tests were repeated and values of 460 meter-ohms, 185 meter-ohms, and 70 meter-ohms, respectively, were obtained. These values are considerably higher than the previous tests in the same general locations. Using newer curves, the lightning damage expectancy was again calculated and found to be 3 cases per 100 miles per year in the first 3-mile section, 1.2 cases per 100 miles per year in the second section, and in the third section negligible. With two *104* copper shield wires, it was calculated that the damage expectancy would be reduced to 0.8 cases per 100 miles per year in the first section and to negligible trouble in the second section. This would be a reduction of 0.066 cases per year in the high-resistivity 3-mile section and a reduction of 0.036 cases per year in the medium-resistivity 3-mile section.

These estimates are based on locating the shield wires approximately one foot apart and one foot above the cable. To place the wires it would be necessary to locate and stake accurately the cable for route and depth. A 2-share plow would then have to be assembled specially for the job.

It was thought that the trouble expectancy figures discussed in the foregoing might be low. In an attempt to verify this, a study was made of the records for the Neillsville–Eau Claire section of the Stevens Point–Minneapolis coaxial cable, which is buried under somewhat similar circumstances. This cable is protected by two *165* copper shield wires. The records showed a damage incidence of 3.1 cases per 100 miles per year. This is 2.6 times as much damage as was estimated using the same method previously mentioned. Assuming that the expectancy figures for the cable under consideration were increased in the same proportion, there would still be an improvement of only 0.17 cases per year for the high-resistivity 3-mile section. It was still felt that this was not sufficient improvement to justify the cost of the shield wires. It was decided to study the matter further and to await results of studies being made by the Bell Telephone Laboratories in the east.

Fig. 4. Second protected terminal north of Wisconsin Rapids looking south

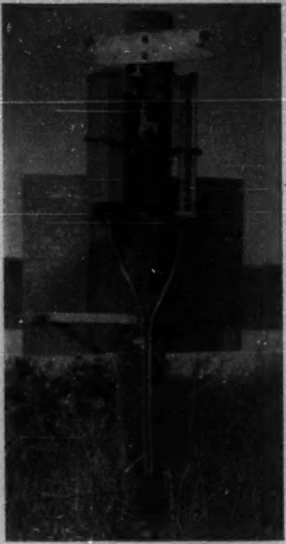


Fig. 5. Third protected terminal north of Wisconsin Rapids looking south

1953 Damage and Remedial Measures

During a storm on July 1, 1953, the cable was again damaged. In this case trouble occurred at four scattered locations in the north mile of the high-resistivity section. Although there was no above-ground evidence of lightning strokes at two of the locations, sheath to conductor burns occurred, indicating a near-by stroke. At the other two locations only conductor to conductor burns were involved.

It was decided that some action now would have to be taken. Two years before, full-count protection, which consists of carbon gaps between every conductor in a cable and the sheath, had been placed on an aerial intercity cable where it crossed a high rocky ridge and had been persistently damaged. In this case, there was a terminal serving a radio station near the middle of the exposed section. Because some conductors were brought out and protected, it was considered desirable to add protection for all other conductors in the cable at this location. Full-count protection provides a path through the carbon gap between any conductor and sheath, and through two gaps in series between any conductor and any other conductor in the cable. Potentials should thus be equalized in the cable without insulation breakdown. This method is normally used on local cables where conductors are frequently brought out and connected to exposed wire or to subscribers' stations. In dealing with intercity cables, it is usually considered more practical to isolate completely the conductors from the sheath, by means of higher breakdown insulation.

The full-count protection had been remarkably successful in eliminating the persistent damage in the case previously mentioned and it was decided to apply it on the Stevens Point–Wisconsin Rapids cable. A 26-pair and a 16-pair protected terminal were placed on a marker post at two locations. One was placed in about the center of the 3-mile section of highest earth resistivity and the other approximately $1^1/_2$ miles to the north. These locations are shown on Fig. 2 with the installations shown in Figs. 3 and 4. These terminals were bridged onto the quads in the cable and provided a 700-volt carbon gap to the sheath on all pairs. This work was completed during the fall of 1953, but before that time two more burns had been found in the high-resistivity earth area. In neither of these had there been a breakdown to the sheath. It is believed that this trouble may have been caused by old burns, possibly from the July storm of that year.

1954 Damage and Extension of Remedial Measures

On July 3, 1954, the buried cable was damaged at eight scattered locations. This damage in general, was centered around the junction with the spur to Rudolph. There was one burn, however, quite near the junction with the aerial cable. The spur to Rudolph consists of a little less than 1/2 mile of open wire, terminating in repeating coils on the terminal pole. Two of the pairs connected at this terminal were involved in a burn to sheath in a sleeve on the terminal pole and in some of the burns in the main cable. Protection at the Rudolph office was not operated and the repeating coils on the pole were not damaged. It appears that the lightning was not brought into the cable from the aerial wire, but by a rise in sheath potential caused by lightning strokes to ground. This damage is located in the area of lowest earth resistivity where estimates of trouble expectancy indicated negligible damage.

In view of this damage in the low-resistivity earth area, it was decided also to place protected terminals on the cable in this area. Terminals as described in the foregoing accordingly were added at a point near the junction with the aerial cable, at the location of the spur to Rudolph, and at a point about midway between this location and that of the northmost terminals placed previously. These locations are shown on Fig. 2 and the installations are shown in Figs. 5 and 6.

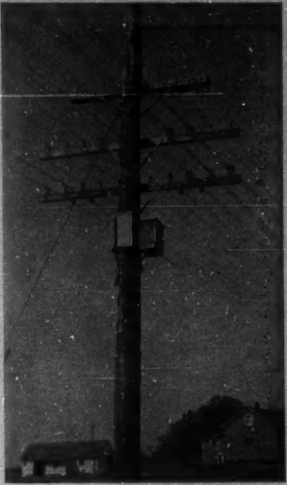


Fig. 6. Terminal pole at spur to Rudolph

At this time a microwave tower was being planned in the area east of Rudolph and the supervisory circuits were to be routed from the same terminal. The full-count protection at this location was provided with the rearrangements necessary to make these circuits available.

With the investigation of this latest damage, two isolated trees were noted which were quite near the cable route. Although these trees showed no visible evidence of having been struck by lightning, it was thought that they constituted a potential hazard and arrangements were made to have them removed.

The installation of the protected terminals and the removal of the trees was completed late in 1954. During 1955 and the early part of the 1956 lightning season, no lightning trouble has been experienced on this cable.

There was considerable concern about the possibility of excessive service interruptions caused by permanent grounding of the carbon gaps in these terminals. The maintenance personnel were instructed to report any such occurrences.

For the period of over 2 years for the first two installations, and over one year for the latter three installations, service interruptions have occurred on only one occasion. On other occasions, after heavy storms, the blocks were examined and found to be heavily dusted.

Conclusions

It is recognized, that this short period is not adequate to indicate fully the value of this method of protection for intercity cables, since longer periods have occurred without damage. It will not be possible completely to evaluate the method in this case, because a new cable along the same route is being planned for installation late in 1956 or early in 1957. This will have a shielding effect on the existing cable. The new cable will have what is called Lepeth sheath. This consists of a layer of extruded polyethylene between the conductors and the lead sheath. Insulation between the sheath and conductors with a dielectric strength of over 20 kv is provided. This type of cable, when carefully spliced to maintain the full dielectric strength through the splices, is very resistant to lightning damage.

The lapsed time is not sufficient to establish definitely the effectiveness of this method of protecting intercity telephone cables, but its consideration for adoption in similar situations appears warranted.

Electronics in Telephone Switching Systems

A. E. JOEL
MEMBER AIEE

DURING the last 40 years a great background and fund of knowledge has developed in the field of telephone switching. Constant improvement in available devices has resulted in an increased scope of application. The field has almost reached a point of perfection as an art and is now rapidly entering a more scientific era.

The tools of the present-day telephone system design engineer are well known; shown in Figs. 1(A) and 1(B) are the relay and the various forms of electromechanical switching apparatus. But over the years, while the art employing these tools was developing, the field of electronics has been developing also. Electronic applications have formerly been most important in relation to sensitivity rather than speed. Even in the telephone switching field, this property of sensitivity has provided better signaling and more accurate timing.

With World War II, however, came a recognition of the speed advantages of electronics. This characteristic was exploited primarily in the quantizing of information in both transmission and information-processing equipment. In the latter field, new digital computers made their appearance and resulted in the development of new forms of electronic devices, most important of which are the "bulk memory" devices.[1]

In the postwar period, the exploitation of another phase of electronics developed from research in semiconductor devices. The transistor is perhaps the best-known invention to emerge from these investigations. The impact of the application of semiconductor devices is yet to be felt in the electronics industry and it will most likely find greatest application in the information-processing field and in communications generally.

Before one can understand and appreciate the impact that electronics will have on the design of new telephone switching systems, it is necessary to consider the question: "What is a telephone switching system?" By evolution it is now generally recognized that the central office portion of a telephone switching system consists of two principal parts, the interconnecting network, or conversation channel, and its control, and certain physical and operational characteristics of these parts; see Fig. 1.

In some switching systems, particularly the progressive direct-control type such as the step-by-step system, these parts are inexorably integrated. In the modern systems, however, they have largely been separated. For purposes of the following discussion, the common-control type of system will be assumed. The bulk nature of the electronic memory devices makes them more readily adaptable to systems of the common control type, where the control functions consisting of the receipt, interpretation, and processing of input signals, and the effecting of output signals, may be concentrated.

Interconnecting Network

In electromechanical switching systems the interconnecting network is composed of crossbar switches or other electromechanical devices. Each connection through the network is physically separated in space from the others and hence this network can be called generically a space-division type. It is subdivided functionally into the concentration stage, where active lines are separated from those not being called or served at a particular time; the distribution stage, where interconnection of active lines and trunks is accomplished; and an expansion stage, where active call paths are connected to selected destinations.

In electronic switching systems, three classes of switching networks have been described[2]:

1. "Space division," similar to the space division for electromechanical apparatus except that electronic devices such as gas tubes are employed in place of mechanical contacts as the crosspoint element.[3]
2. "Time division," where calls are samples in time, each one being given a time slot on a single channel.[4,5]
3. "Frequency division," such as employed in carrier systems where each call is modulated to a different frequency level on a single transmission medium.[6,7]

In electronic switching the interconnecting networks derive their basic characteristics from the known methods of telephone transmission. Since transmission techniques are used it is generally not feasible to pass d-c signals through such networks. Also, certain a-c signals such as 20-cycle current which are now

---

Paper 56-668, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 28, 1956; made available for printing May 8, 1956.

A. E. Joel is with Bell Telephone Laboratories, Inc., Whippany, N. J.

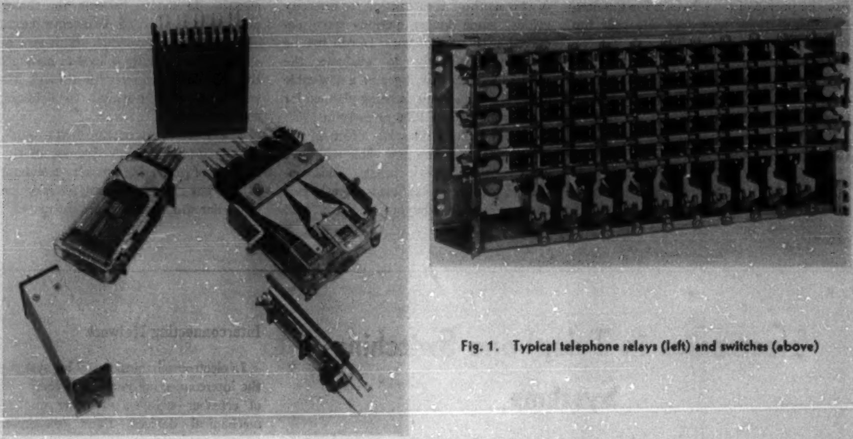


Fig. 1. Typical telephone relays (left) and switches (above)

used for ringing are of a power level so high that they would overload the electronic switching devices employed. It thus seems that to accomplish switching with an electronic interconnecting network, a change is required in the customer's apparatus to make it capable of responding to a lower level alternating current for the call signal. Telephone sets with transistor amplifiers and an acoustical horn are being developed. Interrupted tones in the voice-frequency range can be used effectively to call the user to the telephone.[8]

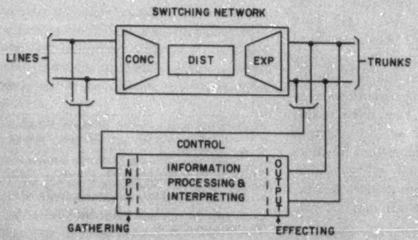
As in most electromechanical switching networks, the concepts of connecting successive stages of switching devices (stages to perform the functions of concentration, distribution, and expansion) to form the network also apply. Since there is more than one method of interconnection, the successive stages of a network may employ either electronic or electromechanical switching techniques, or both. In electromechanical switching, different devices may also be used in different stages.

In electromechanical space-division networks, certain types of crosspoints are better adapted to common-control operation than are others. Systems with electromechanical selector switches most generally are set progressively. In systems with relays or relay-like crosspoints, all crosspoints involved in a connection may be actuated simultaneously. In either case the switching device, or the circuit in which it is used, has a form of memory. This memory, shown as M in Fig. 2, may be the ability of a selector to be held mechanically in a particular path-connecting position or in a locking or holding circuit associated with a crosspoint relay or crossbar switch magnet.

To minimize the time consumed by the common-control elements, simultaneous operation of relay or relay-like crosspoints is most desirable. However, this type of network requires a grid of link testing and control leads such as shown in Fig. 3 for a typical stage of a crossbar switching network. In a network of this type, the calling rate capacity is limited by the slow actuating speed of the electromechanical relay or switch. Efficient network configurations can be devised for large capacity. To set up connections at a high rate in such a network requires a plurality of controls, each capable of operating on all or part of the network. In any case, the controls function in parallel on the network because of the speed considerations.

With electronics applied to space-division switching networks, two improvements over the operation of relay-type space-division networks may be achieved. First, the speed of operation of the crosspoint elements may be made high enough so that only one control is needed to operate on networks of the size now requiring a plurality of controls. Second, the properties of proposed electronic crosspoint elements are such that the principle of "end-marking" may be employed; see Fig. 4.

Here in contrast to the grid of testing and actuating wires required in electromechanical versions of space-division



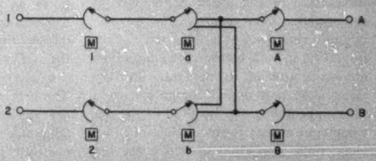


Fig. 2 (left). Principal parts of common-control telephone-switching system

Fig. 3 (above). Space-division switching

Vol. Pt. 1 1956

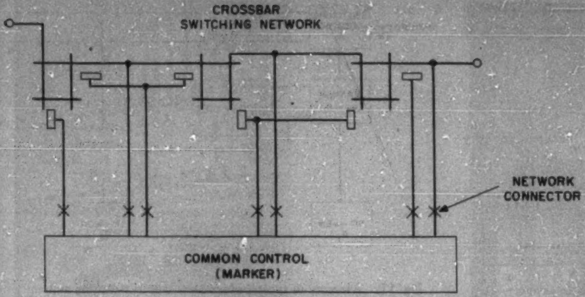

CROSSBAR
SWITCHING NETWORK
NETWORK
CONNECTOR
COMMON CONTROL
(MARKER)

Fig. 4. Typical common control of a crossbar switching network

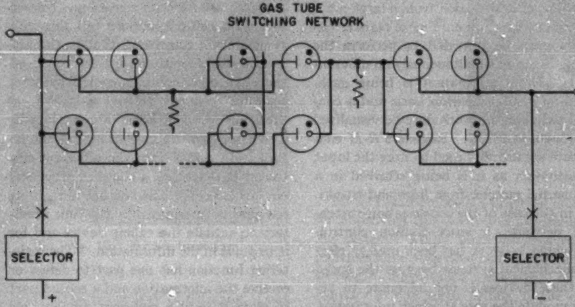

GAS TUBE
SWITCHING NETWORK
SELECTOR
SELECTOR
+
−

Fig. 5. Typical end-marking control of a gas-tube switching network

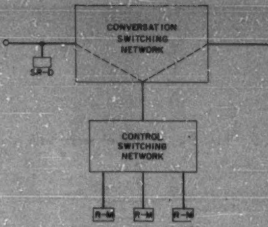

CONVERSATION
SWITCHING
NETWORK
SR-D
CONTROL
SWITCHING
NETWORK
R-M
R-M
R-M

Fig. 8. Control access

networks, the electronic space-division switching network requires only the selectors at each end of a desired network connection to apply the marking potentials. The electronic crosspoint element will be actuated if the link to which it connects is idle. Eventually all available paths between input and output will be marked. As means must be provided for sustaining only one of the possible idle paths, the memory property of the crosspoint device takes over to hold the path until it is released by release marks or removal of the sustaining voltages. In space-division networks, the memory requirements must be satisfied just as in electromechanical networks.

Multiplexing and carrier transmission systems[9] employ time and frequency division but the physical terminals at both ends of a channel for which the facilities are derived have a 1-to-1 correspondence which can only be changed manually. In a switching system, the input-output relations must change automatically as required for each call. Here the need arises for a changeable memory for associating a given time or frequency slot to a particular call at any given time. At some other time, these points in time or frequency must be capable of being assigned automatically to different inputs and outputs. For the period that they are assigned, some form of memory must record this assignment and this memory is consulted continuously or periodically for the duration of the call.

With time-division switching, this new concept in the use of memory in a switching network appears most clearly; see Fig. 5. To associate an input with an output during a time slot, the memory must be consulted which associates the particular input with the particular output. To effect the connection during a time slot the input and output must be selected. A memory is consulted to operate high-speed selectors simultaneously for both the input and output. Each selector receives information from a memory which actuates crosspoints to associate the input or output with the common transmission medium. The information from the memory which controls the selection process is known as an "address." The crosspoint is nonlocking since it must open when the selector re-

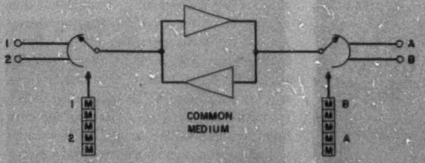

1
2
A
B
1
M
M
M
M
M
2
COMMON
MEDIUM
M
M
M
M
M
B
A

Fig. 6 (above). Time-division switching

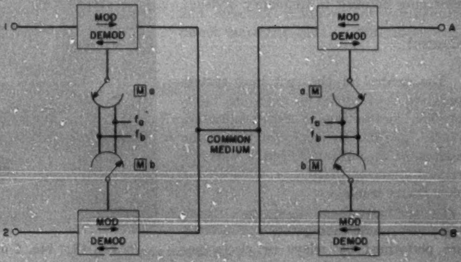

1
MOD
DEMOD
M a
fa
fb
M b
2
MOD
DEMOD
COMMON
MEDIUM
MOD
DEMOD
A
a M
fa
fb
b M
MOD
DEMOD
B

Fig. 7 (right). Frequency-division switching

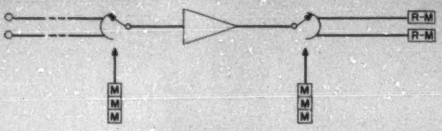


Fig. 9. Time-division control access with separate functional memory

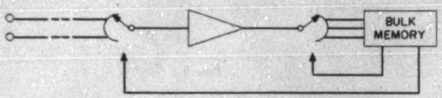


Fig. 10. Time-division control access with bulk memory

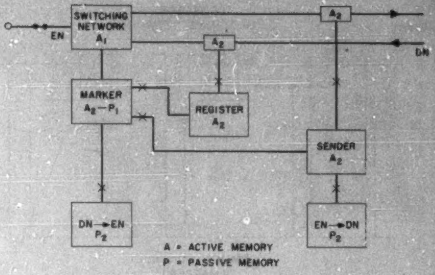


Fig. 11. Memory in typical electromechanical switching system

ceives its next address. The individual memory of crosspoints for space-division networks has thus been changed by time division to changeable memory, usually in the form of a coded address associated with each time slot. Furthermore, since the successive addresses actuate the same selectors and hence may be held in a common high-speed device, electronic bulk memory is ideally suited for this task. The memory must be changeable to allow for different associations of input to output at different times.

In frequency division, the control characteristics of the interconnecting network require a modulation frequency to be assigned to each simultaneous conversation applied within the bandwidth of the common medium. As shown in Fig. 6, the application of the modulation frequencies requires a separate selector for each input and output. These selectors are nothing more than space-division switching networks and therefore require memory in the switching devices whether they are electromechanical or electronic.

In addition to memory for associations within the switching network, selecting means are also needed to activate a terminal to be chosen in space division (e.g., Fig. 4), to place address information in the proper time slot in time-division switching, or to set the frequency-apply-switching network in frequency division.

Control

The control of the switching system provides the facilities for receiving, interpreting, and acting on the information placed into it. In particular this is the address of the output desired. A service request detector (SR-D) is provided for each line or trunk.

In electromechanical systems these logic- and information-gathering functions are performed by relays or electromechanical switches. In order to keep up with the flow of information from a large number of customers, a number of register circuits must be provided to perform the same function simultaneously on different calls. Here information is being gathered on a space-division basis and a control switching network may be visualized as shown in Fig. 7. Registers *R-M* constitute the memory used to store the input information as it is being received in a sequential manner from lines and trunks. As in the case of the conversation-switching network, a space-division control-switching network has been used in electromechanical systems because the speed of these devices is not adequate to accommodate the rate at which information flows into the system. In the step-by-step system the control and conversation switching networks are coincident. In the no. 5 crossbar system[10] the same network is used for both control and conversation on call originations but the functions are not coincident, that is, the network is used for either control or conversation. In other common-control systems, separate control networks known as "register or sender links" are employed.

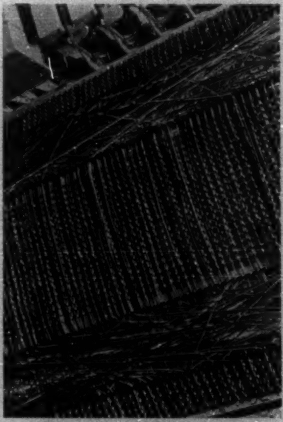


Fig. 12. No. 5 number group

When relays are used to receive the information pulsed into the office by customers or operators, a plurality of register circuits is needed. The number of registers required is determined by the time necessary to actuate the calling device and for it to pulse in the information. The registering function has one part to detect or receive the information and a second part to store it until a sufficient amount has been received for processing. The processing function is usually allotted to other circuits such as the markers in crossbar systems.

Since the input of information to a switching system is usually limited to two conductors, a serial form of signaling is used. It would seem only natural that if a detector were fast enough, it could function to receive the serial information on several simultaneously active inputs. Relays are not fast enough to accomplish this, but high-speed time-sharing electronic devices have been designed to perform this information-gathering function. Since it is a time-sharing arrangement it is analogous to the time-division switching. A time-division control access as shown in Fig. 8 requires memory to control the time-division switching functions. Time sharing as applied to the gathering of information in telephone switching systems has been called "scanning." The individual register memories are still in parallel form because of the relatively long time required for sufficient information to be received before processing may start. Higher speed means for placing information into switching systems, such as pre-

Fig. 13. AMA translator

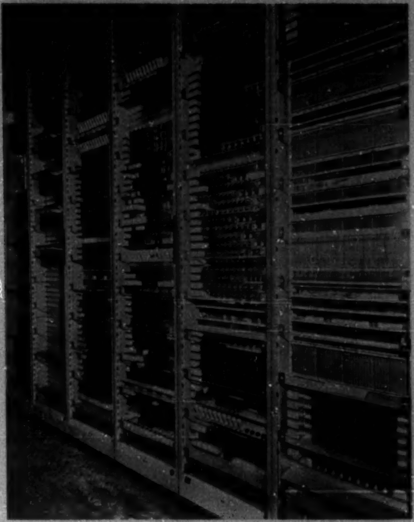


Fig. 14. No. 5 route relay frame

set keysets, is one way of reducing, if not eliminating, this need for parallel register storage in the switching system prior to processing. However, with this type of device one merely transfers the location of the storage from the central office to the customer's telephone set. The fundamental limitation is the rate at which a human being is able to transfer information from his brain into some physical representation. A time-division control access with separate functional memory is shown in Fig. 9.

Fig. 15. No. 4A card translator

Lower cost memory is a practical means for improving this portion of the switching system.[9] Many small low-cost relay registers have been designed and placed in service.[10] Electronics, however, offers memory at one tenth, or less, of the cost per bit if used in large quantities with a common-memory-access control. New low-cost bulk electronic memories are now available to be used in this manner. As shown in Fig. 10, the memory for the control of the time-division control access network and the register memory may be combined in the bulk memory.

Memory appears in the control portion of a switching system in many ways, some obvious and others more subtle. Fig. 11 shows a typical electromechanical switching system, much like no. 5 crossbar. The active memory, designated A, includes call information storage A_2 whether in a register, sender, or marker during processing. There is also certain pertinent call information storage associated with trunk circuits such as a no-charge class on outgoing calls or the ringing code used in incoming calls. Another type of active memory A_1 has been mentioned in connection with switching networks to remember the input-output associations. In most electromechanical systems active memroy has been implemented with relays or switches.

Another form of memory is also employed in all telephone switching systems and much effort has been devoted to devising improved means for effecting a memory that is not changed with each call but is of a more permanent nature. Examples of this passive memory, P in Fig. 11, are the translations required in common-control systems to obtain certain flexibility between the assignment of lines to the switching network and their directory listing. These translations between equipment numbers (network location) and directory numbers are required to direct incoming calls to the proper ter-

Fig. 16. Wiring side of relay unit

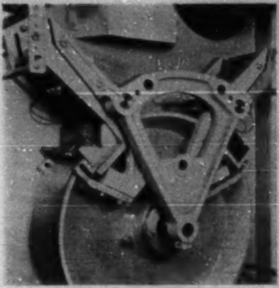


Fig. 17. Magnetic drum

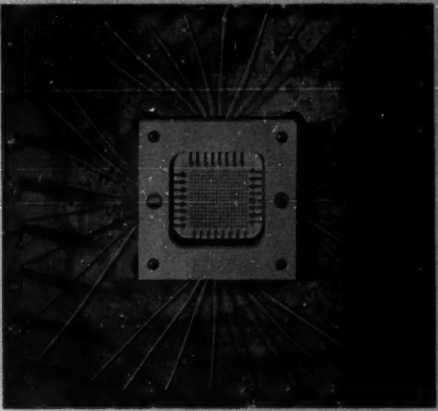
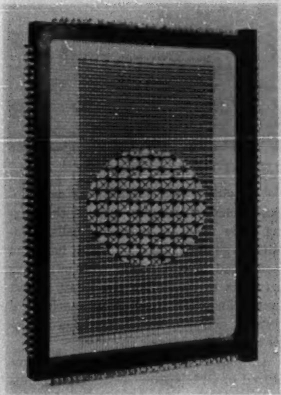


Fig. 20. Ferroelectric array



Courtesy of International Business Machines Corporation

Fig. 18. Magnetic-core array

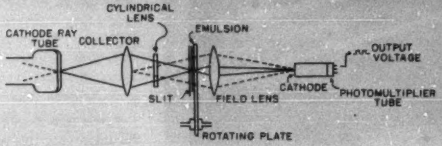


Fig. 21. Photographic storage

from *Proceedings*, Institute of Radio Engineers, October 1953

FLYING SPOT SCANNING A ROTATING DISC

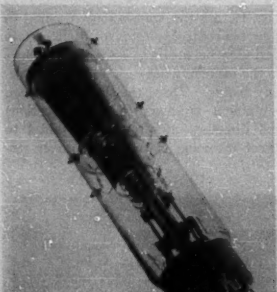


Fig. 19. Electrostatic storage tube

minals, such as the number group frame in no. *5* crossbar of Fig. 12, and to provide information for charging purposes on originating calls, such as the AMA "dimond" ring translator[11] of Fig. 13. Each of these translators for a 10,000-line office represents about 10^6 bits of information. Passive memory is also used to translate central office codes into routing information. In local central offices this is also done by cross-connections as shown in Fig. 14.

Another form of passive memory is the punched card or tape used in telephone accounting systems. A step toward electronic memory is the card translator which provides routing information in the crossbar toll switching system,[12] see Fig. 15. Here the cards represent passive memory and are selected and read by a combination of electromechanical action and light-beam sensing with phototransistor detectors. One such device equipped with 1,000 cards represents the storage of approximately 10^6 bits of information.

In all of these types of passive memory, speed limitations are involved in the choice of devices used within the memory or the access to it. This is one reason why these translators are subdivided so that the various portions may be used in parallel in order to satisfy the total information-processing needs of the office.

A discussion of passive memory would not be complete without one further illustration. Fig. 16 is a wiring side view of a typical relay circuit in the information-processing portion of a switching system. It could be any other unit, such as a trunk circuit, but in any case each wire on such a unit is remembering some passive relationship between the active portions of the circuit, such as relays. This is the memory of the contact and coil interrelationships as conceived by the designer and based on the requirements of what the circuit is required to accomplish. It is the program of what the central office must do at each step of every type of call. Modern digital computers have been built with the ability to store programs in bulk memories for the solutions of the various types of problems put to them. It is conceivable that the program of a telephone central office may also be stored in bulk memories to eliminate the need for much of the fixed wiring such as appears in relay call-processing circuits.

The form of memory available in electronics is considerably different from that which has been previously available.

Electronic memory has been characterized as "common-medium" or "bulk" memory. Here, a single device is capable of storing more than a single bit of information, which is the limit of most relays or other devices capable of operating in a bistable manner. A number of different types of electronic bulk memories have been devised for digital processing. They differ appreciably in physical form, each taking advantage of the phenomenon of some different area of the physical sciences—electrostatic, electromagnetic, or optic. Magnetic tapes[13] and drums[14] (Fig. 17), cores[15] (Fig. 18), electrostatic storage in tubes[16,17] (Fig. 19) and ferroelectrics[18,19] (Fig. 20), and photographic storage[20] (Fig. 21) are available.

Several properties of these memory devices are of interest, the primary being the speed with which stored information may be read. This is known as "access speed." Another property of these common-medium memory systems or devices is the ability to change what has been written. If the changes can be made rapidly enough, they may be used in electronic systems in much the same manner as relays are used in electromechanical systems to process information. If the change must be made relatively infrequently, such as changing photographic plates, they may be used as substitutes for the type of memory in these systems which are provided by cross-connections and wiring. The required fixed or semipermanent electronic memory may be characterized primarily by a high reading speed, large capacity, and the ability to hold stored information even during prolonged intervals of loss of power. The amount of memory is measured in terms of binary digits or "bits." The number of bits equivalent to a single cross-connection can be rather large. Therefore, an electronic memory replacing a fixed memory such as in the card translator in modern electromechanical systems should be high in bit capacity, say from 10^6 to 10^7 bits for 10,000 lines.

One way in which electronic memory for various system applications may be evaluated is given in Fig. 22. This chart shows, for the various forms of storage, the relation between the capacity in bits and cycle time, which includes access, reading, and, if necessary, the regeneration time of the stored information. For the sake of simplicity, ferroelectric and magnetic-core memories have been combined as co-ordinate access arrays. Single-bit electronic memory will be described in more detail later.

In the control portion of a switching system it is not only necessary to gather

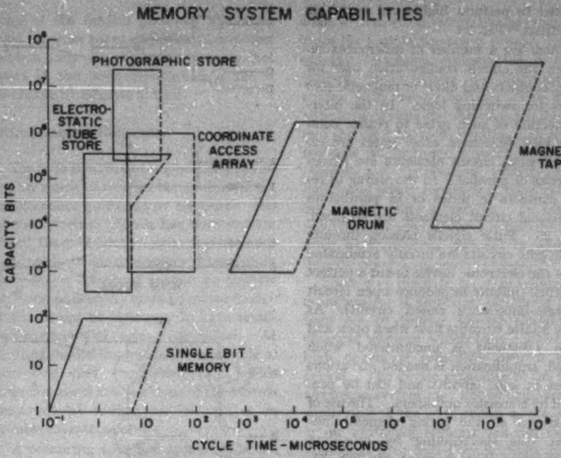


Fig. 22. Memory system capabilities

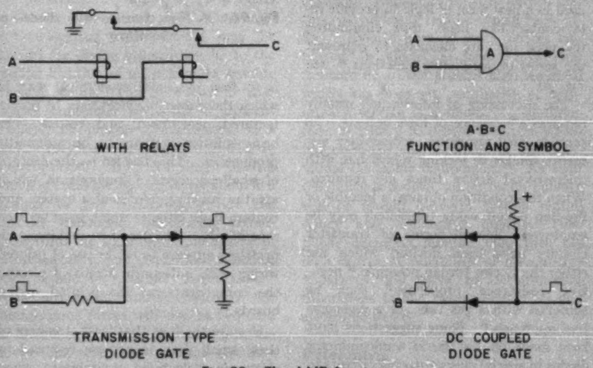


Fig. 23. The AND function

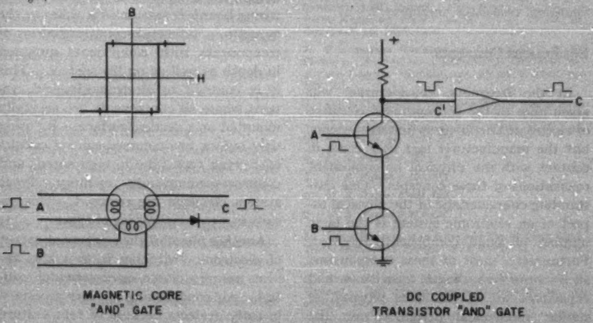


Fig. 24. Other AND circuits

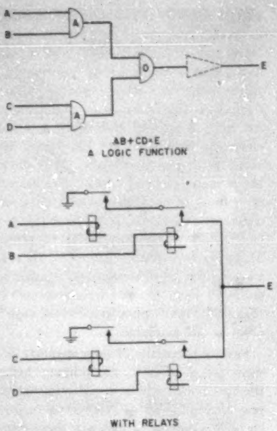


Fig. 25. A logic function with relays

and store information, but the information must be interpreted and appropriate action taken. This function is called "processing." Processing circuits control the information-gathering and storage functions and perform logical functions to produce the necessary flow of information. In the logic circuits of electronic systems, to keep pace with the time-sharing nature of the information-gathering function, the devices used must be several orders of magnitude faster than their counterparts, the relays, of the electromechanical system. The scanning and bulk memory access speeds must be comparable in speed if they are not to create a bottleneck. All portions of the system must be in balance in regard to time.

Devices and techniques for use in the design of high-speed logic circuits are available,[21] and information processing previously carried out by complex relay circuitry may be carried out in microseconds instead of milliseconds. Devices such as semiconductor diodes and transistors seem to be pointing the way to the future in performing these functions.[22] Previously, hot cathode tubes with high power consumption were needed to achieve the same functions at similar high speeds and for a long time this has been one of the greatest deterrents to electronic switching.

Semiconductor diode gate circuits are now quite familiar[23] and take the place of the conventional make-and-break contacts in the electromechanical switching art (see Fig. 23 for the AND function). Magnetic-core circuitry is also being exploited to perform high-speed switching functions,[24] Fig. 24.

There are a number of differences between the circuit configuration used for relay contacts and diode or magnetic-core gates for switching logic. In the interconnection of such gates to realize complex logic functions, other gates are required when circuit elements are placed in series or parallel. In the wiring of relay contacts in series or in parallel no additional circuit elements are required, Fig. 25. Pulse signals passing through diode gate circuits are usually attenuated since the electronic device is not a perfect switcher (infinite-impedance open circuit to zero-impedance closed circuit). As some minite currents flow when open and some resistance is encountered when closed, amplification is needed at various places in logic circuits and can be provided by transistor amplifiers. The use of transistors as the gating element eliminates this shortcoming by providing amplification in each gate; see Fig. 24. Transistors have also been successfully used in a new form of logic to provide relay-contact-like logic, thus eliminating the need for gate elements to represent the series of paralleling functions;[25] see Fig. 26.

The processing of information usually requires a sequence of logic actions. To provide such sequences, momentary elements similar to locking relays but with microsecond action times are required. When this condition obtains, a bistable or flip-flop circuit using transistors may be employed. Several forms of transistor circuits have been devised which use either the Eccles-Jordan principle,[26] negative resistance properties[27] such as achieved with a gas tube, or a regenerative approach.[28] Some suggestions have been made on the use of semiconductor diodes in special energy-storing circuits to amplify pulses instead of the more conventional transistor amplifiers.[29]

Equipment Concepts

In the foregoing, consideration was given only to the concepts and circuitry of electronic telephone switching systems, but the manufacturer and user come in contact with the physical or equipment realizations of these concepts. One outstanding characteristic of the physical aspect of an electronic system is the large number of small components required. Fortunately, most of these components, such as resistors, diodes, transistors, and capacitors, are of the same physical or similar mechanical design. From the manufacturer's point of view, the problem is to find the most economical way in which these many devices may be manufactured, assembled, and tested. The basic solution appears to be automatic production. This has led to the concept of small packages of components, which are the building blocks of a system and contain basic circuits which may be used repetitively. The trend in making such packages appears to be the use of printed wiring with automatic means of placing the components on the printed wiring boards.[30]

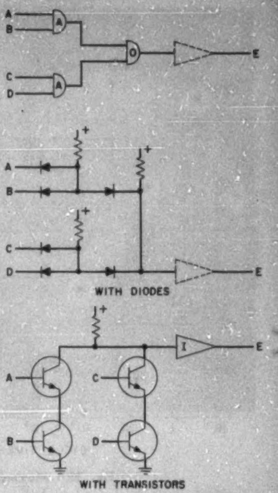


Fig. 26. A logic function with diodes or transistors

Despite the fact that large numbers of these small components are required in electronic telephone switching, when equipment using a multiplicity of printed wiring boards is assembled it takes on the aspect of a 3-dimensional arrangement of components with components mounted in depth as well as on the surface. This is in contrast to electromechanical systems where all components are generally mounted on a vertical surface. By using only one or two common-control circuits of a given type (due to high speed) and common-medium-bulk memory, fewer system elements are required. This results in a material saving of space.

Another phase of the equipment aspect of electronic switching is that the devices require closer environmental control. Air conditioning appears necessary in early systems because of temperature limitations and other characteristics of

Vol.
Pt. 1
195

some of the devices presently available. Also, vacuum tubes and other high-power devices may develop objectionable hot spots in the equipment which make it advisable to exhaust hot air.

Maintenance Concepts

There is insufficient experience at present on which to base an estimation of the maintenance problems of electronic telephone systems. Much has been written about the problems encountered in maintaining electronic computers; however, in designing a telephone system an entirely different philosophy must be pursued since it should not be necessary to have maintenance forces of engineering caliber. At no time should the system be incapable of accepting and completing calls. This does not mean that portions of the system may not be worked on for routine or trouble maintenance.

A promising approach appears to be the use of marginal condition routine tests for detecting in advance components which are about to fail.[31] Automatic trouble-locating arrangements may be devised for giving information as to the specific location of a package in trouble when it occurs. This automatic trouble locator combined with the equipment concept of plug-in units means that service may be maintained without long interruptions. By designing devices which are reliable, by employing them in a manner to give maximum service life, and by judiciously introducing redundancy into the equipment, the chance of simultaneous failure of any two identical parts should be extremely improbable.[32] With automatic trouble locating, the maintenance forces will not be required to have a thorough understanding of the device characteristics and the circuitry used. Centralized repair of defective units as in modern telephone transmission systems,[33] and perhaps even expendability of defective units, are a distinct possibility.

As a result of some of these maintenance considerations, it is quite likely that equipment in the future, in addition to being smaller and more compact, will appear more generally in enclosed low cabinets rather than in exposed frames. The administrative control may be from consoles rather than from vertical panels. More attention will be paid to appearance. The appurtenances, such as ladders required for high frames in electromechanical systems, may be eliminated.

Another change in concept which may come with electronics in telephone switching is the form of the power supply. Present-day telephone systems use a centralized single-voltage d-c distribution system with reserve battery. The wide variety of devices and associated voltages, and the need for close regulation in some portions of electronic systems, makes a reliable a-c distribution system with individual power rectifiers at the point of use appear quite attractive. To insure reliability of service, the a-c distribution must be continuous and not dependent directly upon commercial sources.

There is no question but that reliability is imperative if electronic switching systems are to survive among electromechanical systems which have achieved a high degree of reliability over a long period. The device reliability of the first electronic system may not be comparable since some of the components of the electronic switching systems will not be as reliable in their initial applications as the least reliable component in present-day systems. Reliability must be earned and this will probably require considerable effort. Even if initially some devices employed in electronic systems do not measure up to the present high standard which has been set, continuity of high-quality service is a must. It is, therefore, necessary to design a system which will mask the shortcomings of any individual electronic component.[32] As their reliability is proved, an optimum balance will be sought between system redundancy and component quality. Telephone engineers familiar only with the high degree of reliability of present-day apparatus will have to accommodate themselves to the characteristics of new electronic devices.

References

1. A Survey of Digital Computer Memory Systems, J. R. Eckert. *Proceedings*, Institute of Radio Engineers, New York, N. Y., v[illegible]1, no. 10, Oct. 1953, pp. 1393–1406.

2. Electronic Telephone Exchanges, T. H. Flowers. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. I, 1952, pp. 181–201.

3. Cold Cathode Tubes for Audio Frequency Signaling, W. A. Dopp, M. A. Townsend. *Bell System Technical Journal*, New York, N. Y., no. 32, Nov. 1953, pp. 1371–91.

4. *U. S. Patent 2,387,018*.

5. *U. S. Patent 2,490,833*

6. *U. S. Patent 2,408,462*

7. *U. S. Patent 2,379,221*.

8. Tone Ringer May Replace Telephone Bell. *Bell Laboratories Record*, New York, N. Y., Mar. 1956, pp. 116–17

9. Time Division Multiplex Systems, W. R. Bennett. *Bell System Technical Journal*, no. 20, 1941, p. 199.

10. The Number 5 Crossbar Dial Telephone Switching System, F. A. Korn, James G. Ferguson. *Electrical Engineering*, vol. 69, Aug. 1950, pp. 679–84.

11. Number 5 Crossbar AMA Translator, T. L. Dimond. *Bell Laboratories Record*, Feb. 1951, p. 62.

12. The Card Translator for Nationwide Dialing, L. N. Hampton, J. B. Newsom. *Bell System Technical Journal*, no. 32, Sept. 1953, pp. 1037–98.

13. Review of Input and Output Equipment Used in Computing Systems. *AIEE Special Publication S53*, Mar. 1953.

14. Magnetic Drum for Digital Information Processing Systems, Cohen, A. A. *Mathematical Aids to Computation*, vol. 4, Jan. 1950, pp. 31–39.

15. Multidimensional Magnetic Memory Selection System, M. K. Haynes. *Transactions*, Professional Group on Electronic Computers, Institute of Radio Engineers, New York, N. Y., Dec. 1952, pp. 25–29.

16. Digital Memory in Barrier Grid Storage Tubes, M. E. Hines, M. Chruney, J. A. McCarthy. *Bell System Technical Journal*, no. 48, Nov. 1955, p. 1241.

17. Storage Tubes and Their Basic Principles (book), M. Knoll, B. Kazan. John Wiley & Sons, Inc., New York, N. Y., 1952.

18. Ferroelectrics for Digital Information Storage and Switching, D. A. Buck. *Report R212*, Digital Computer Laboratories, Massachusetts Institute of Technology, Cambridge, Mass., June 1952.

19. Ferroelectric Materials as Storage Elements for Digital Computers and Switching Systems, J. R. Anderson. *AIEE Transactions*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 395–301.

20. Photographic Techniques for Information Storage, G. W. King, G. W. Brown, L. N. Ridenour. *Proceedings*, Institute of Radio Engineers, Oct. 1953, pp. 1421–28.

21. Synthesis of Electronic Computing and Control Circuits (book), Staff of Harvard Computation Laboratory. Annals of Harvard Computation Laboratory, vol 27, Harvard University, Cambridge, Mass., 1951.

22. Typical Block Diagrams for Transistor Digital Computor, J. H. Felker. *AIEE Transactions*, vol. 71, pt. I, July 1952, pp. 175–82.

23. Engineering Multistage Diode Logic Circuits, B. J. Yokelson, W. Ulrich. *Ibid.*, vol. 74, pt. I, Sept. 1955, pp. 466–65.

24. Pulse Switching Circuits Using Magnetic Cores, M. Karnaugh. *Proceedings*, Institute of Radio Engineers, vol. 43, no. 5, May 1955, pp. 570–84.

25. Surface Barrier Transistor Switching Circuits, R. H. Beter, W. E. Bradley, R. B. Brown, M. Rubinoff. *IRE Convention Record*, Institute of Radio Engineers, pt. 4, 1955, pp. 139–145.

26. Two Transistor Binary Counter, R. L. Trent. *Electronics*, New York, N. Y., vol. 25, July 1952, pp. 100–01.

27. Transistors in Switching Circuits, A. E. Anderson. *Proceedings*, Institute of Radio Engineers, vol. 40, no. 11, Nov. 1952, pp. 1541–58.

28. Regenerative Amplifier for Digital Computer Applications, J. H. Felker. *Ibid.*, pp. 1584–96.

29. Promising Electronic Components—Diode Amplifiers, *Radio Electronics*, New York, N. Y., Nov. 1954, p. 45.

30. Mass Production of Electronic Subassemblies, A. A. Lawson. *Electrical Manufacturing*, New York, N. Y., vol. 54, Oct., 1954, p. 134.

31. Diagnostic Programs and Marginal Checking in Whirlwind I Computer, N. L. Daggett, E. S. Rich. *IRE Convention Record*, Institute of Radio Engineers, pt. 7, 1953, pp. 48–54.

32. Increasing Reliability by the Use of Redundant Circuits, C. J. Creveling. *Proceedings*, Institute of Radio Engineers, Apr. 1956, pp. 509–15.

33. Servicing Center for Short-Haul Carrier Telephone Systems, A. L. Bonner. *AIEE Transactions*, vol. 3, pt. I, Sept. 1954, pp 388–96.

(*Discussion follows.*)

Discussion

W. H. T. Holden (Pasadena, Calif.): The discussion of common-medium switching systems in this paper appears to be based on the studies made on such systems in the period just preceding World War II. Both time- and frequency-division systems were considered. While Mr. Joel has shown the 2-wire repeater found necessary in these systems in the common medium, the requirements for this were not mentioned in the paper.

It was found that in any common-medium system there is an inherent loss in transmission, due both to impedance mismatch and to the energy absorbed by other terminals in frequency division, and to the necessity of spreading the energy in the sampling interval of time division systems over the entire interval between pulses. This results in substantial losses, reaching as much as 72 decibels in the case of a 2,000-by-2,000 frequency-division array.

Gains of this magnitude are a difficult problem in a 22-type repeater in the range of frequencies around 10 megacycles. The power level needed to supply a level of −6 to the output side is also a problem.

While it might have been possible in theory to supply gain in the individual channels, this would have made the system economically impossible with vacuum-tube amplifiers. It may well be that in the course of time transistor amplifiers will be reduced in cost to the point where this is practicable.

It may also be pointed out that the basic features of switching systems do not seem altered by the introduction of electronic methods. In fact, some system features once used and abandoned may well warrant revival with the better tools available today. The station sender of the Lorimer system is an example of such a device.

A. E. Joel: The two forms of common-media systems together with space-division systems now appear basic to electronic switching systems. No attempt has been made in this paper to evaluate or compare these types. Crosstalk in electronic switching networks as well as inherent losses may impose gain limitations.

The basic differences which occur in electronic switching systems over electromechanical systems is further concentration of the functions. For example, in existing common-control systems per-call memory may be found in locking relays located in such widely dispersed physical requirements as the switching network, trunk circuits, registers, senders, and markers. With the bulk electronic memory, many storage functions associated with calls in progress may be retained in the same compact physical device. The basic requirements for service and the functions which switching systems perform to carry them out have not changed.

Field Experience with the A2A Video System

RAYMOND W. GAST
NONMEMBER AIEE

A NEW video transmission system known as *A2A* has recently been developed for use on local wire channels in television network service. The *A2A* system provides video transmission for frequencies up to 4.5 mc (megacycles) and to distances of approximately 10 miles over balanced cable pairs designed for the purpose. This paper will describe the system from a field point of view, showing typical applications and performance and operating considerations.

The first *A2A* system was installed for field trial in July 1954 and underwent tests for a period of 10 months. This was prototype equipment. Commercial service in New York, N.Y., was initiated with production-type equipment in September 1955 on one studio-transmitter line, followed by eight network-connecting lines in October and November of the same year. Since then, substantial numbers of *A2A* systems have been installed in various cities to provide for growth and to facilitate maintenance of critical lines.

The local video channel forms an important part of the closed-circuit network as used for television broadcasting. *A2A* design is predicated on meeting high-quality performance objectives for an over-all 4,000-mile network which may comprise a number of local circuits as well as intercity facilities.

In the early days of television broadcasting when each station provided its own program material, a signal might originate at a field pickup point, be transmitted to the studio over a short link, and from the studio to the broadcast transmitter over another; these two links would constitute the entire closed-circuit path. With nation-wide networks, circuits have been extended to thousands of miles in length with complex intercity and local facilities operating in tandem. The transmission quality expected on a coast-to-coast circuit excels that which may have been acceptable on the local loops alone a few years ago. Introduction of color has added further exacting transmission requirements to those for black and white. Local video channels constitute only a small part of the network circuit mileage but they form important and critical links in the over-all system. Fig. 1 shows the circuits for a typical network telecast in simplified form. In this particular instance, six local circuits are connected in tandem with the intercity facilities. Most network programs will involve at least three local circuits at the sending end and two at the receiving end.

The makeup of a television network varies from hour to hour according to program demands, and it is not possible to line up or equalize the network on an over-all basis. Each link must be capable of a transmission quality such that when all the necessary links are connected in tandem, the picture will not be appreciably degraded.

Design Considerations

The *A2A* system is designed to use the shielded cable pairs which have been placed extensively for video use in the telephone plant since 1947. Each balanced video pair consists of two 16-gauge wires, polyethylene-insulated and enclosed in a double shield of copper tape. Several versions of video pair have been manufactured which differ slightly in construction but are similar in electrical characteristics. The *A2A* system may

---

Paper 56-972, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted June 8, 1956; made available for printing August 1, 1956.

RAYMOND W. GAST is with the New York Telephone Company, New York, N. Y.

The many persons who participated in the installation and testing of the first *A2A* systems and contributed valuable suggestions include associates of the author in the New York Telephone Company, the Bell Telephone Laboratories, and the American Telephone and Telegraph Company; also H. C. Gronberg, National Broadcasting Company, and W. F. Whalley, Columbia Broadcasting System, who participated in many tests.

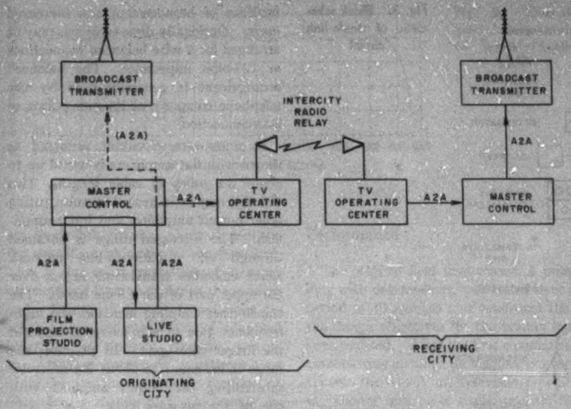

BROADCAST TRANSMITTER
(A2A)
INTERCITY RADIO RELAY
MASTER CONTROL
A2A
TV OPERATING CENTER
A2A
A2A
A2A
FILM PROJECTION STUDIO
LIVE STUDIO
ORIGINATING CITY
TV OPERATING CENTER
A2A
MASTER CONTROL
A2A
BROADCAST TRANSMITTER
RECEIVING CITY

Fig. 1. Connections for typical network telecast

employ any of these shielded video pairs but it is not designed for use with ordinary paper-insulated telephone pairs which have been employed for video to a limited extent in the past.

The circuit lengths over which *A2A* must operate depend on the physical locations of studios and other terminating points, also the locations of telephone buildings which can accommodate repeaters and the available cable routes. In a city, an erratic pattern of terminal and repeater locations is to be expected as illustrated in Fig. 2. The squares represent large studios, master controls, broadcast transmitters, and the operating center for out-of-town networks. There are numerous other studios and pickup points not shown. The circles represent telephone buildings which are suitable for repeater points. The lines indicate principal routes for video cable. Terminal locations are fixed by service demands but there is some choice of routes and of repeater points. Where feasible parallel circuits are assigned in different routes or different cables for added protection of the service.

The circuit lengths range from a fraction of mile to about 10 miles with the majority 1 or 2 miles long. In a large active television center such as New York, local circuits are rearranged, connected, and disconnected with surprising frequency to meet the day-to-day needs of broadcasters and industrial users. Studios, many of which are in leased buildings, may be moved and sometimes change hands from one broadcasting company to another. In addition, there are unpredictable demands for one-time pickup and closed-circuit networks. These requirements are anticipated in the *A2A* system by a variety of amplifiers and cable equalizers which may be assembled in a number of combinations. Portable equipment is available for temporary installations. Both the portable and the rack-mounted equipment is arranged for flexible assignment with cable sections of any length up to the working limit.

Fig. 3 shows a simplified block diagram of a single-link *A2A* circuit where there is approximately 1 mile of cable between the transmitting terminal and the receiving terminal. The circuit is of unity gain with an input of 1 volt peak-to-peak for the composite video signal. The transmitting terminal consists of a cable equalizer which introduces low-frequency loss complementary to the high-frequency loss of a portion of the cable. This is followed by an output amplifier which supplies gain and converts the 75-ohm unbalanced input to a balanced output, the impedance of which matches closely the characteristic impedance of the cable pair. The transmitting terminal has several functions. It effectively isolates the cable pair from the studio equipment and provides a nearly perfect termination for the input signal. Thus it is possible to monitor the signal at the input terminal by means of a very-high-impedance monitoring probe. Secondly, the pre-equalization at the transmitting end of the circuit causes high-frequency components to be transmitted at high level to overcome the cable attenuation losses while the low-frequency components, which are attenuated a lesser amount by the cable, are

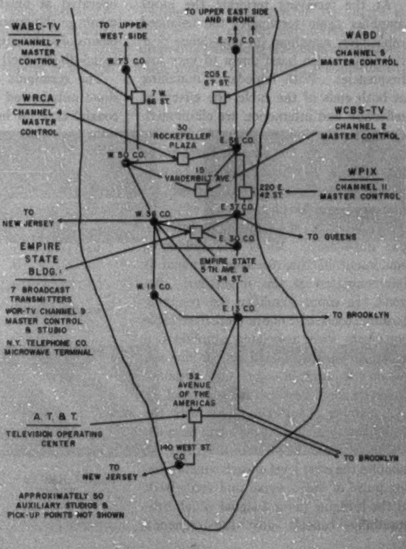

WABC-TV
CHANNEL 7 MASTER CONTROL
TO UPPER WEST SIDE
TO UPPER EAST SIDE AND BRONX
E. 79 C.O.
WABD
CHANNEL 5 MASTER CONTROL
W. 73 C.O.
205 E. 67 ST.
7 W. 66 ST.
WRCA
CHANNEL 4 MASTER CONTROL
WCBS-TV
CHANNEL 2 MASTER CONTROL
30 ROCKEFELLER PLAZA
E. 58 C.O.
W. 50 C.O.
15 VANDERBILT AVE.
WPIX
CHANNEL 11 MASTER CONTROL
220 E. 42 ST.
E. 37 C.O.
TO NEW JERSEY
W. 36 C.O.
TO QUEENS
E. 30 C.O.
EMPIRE STATE BLDG.
7 BROADCAST TRANSMITTERS
WOR-TV CHANNEL 9 MASTER CONTROL & STUDIO
N.Y. TELEPHONE CO. MICROWAVE TERMINAL
EMPIRE STATE 5TH AVE. & 34 ST.
W. 18 C.O.
E. 13 C.O.
TO BROOKLYN
32 AVENUE OF THE AMERICAS
A. T. & T.
TELEVISION OPERATING CENTER
TO BROOKLYN
140 WEST ST. C.O.
TO NEW JERSEY
APPROXIMATELY 50 AUXILIARY STUDIOS & PICK-UP POINTS NOT SHOWN

Fig. 2. Principal video cables and equipment locations in Manhattan

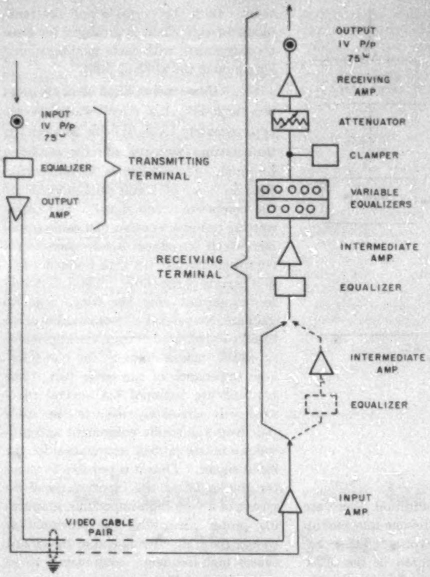

OUTPUT
1V P/P
75Ω
RECEIVING AMP
ATTENUATOR
CLAMPER
VARIABLE EQUALIZERS
INTERMEDIATE AMP
RECEIVING TERMINAL
EQUALIZER
INTERMEDIATE AMP
EQUALIZER
INPUT AMP
INPUT
1V P/P
75Ω
EQUALIZER
TRANSMITTING TERMINAL
OUTPUT AMP
VIDEO CABLE PAIR

Fig. 3. Block schematic of single-link circuit

reduced to a lower level to prevent excessive modulation or overloading in the amplifiers.

At the receiving terminal the input amplifier again terminates the balanced cable pair, matching its impedance over the entire video band down to very low frequencies. With impedance matching at both ends of the cable, the effects of reflection and interaction are eliminated. Consequently, it is necessary to equalize only for the true loss characteristic of the line and the equalizer units can be provided in various sizes to correct for measured lengths of cable.

The input amplifier has a 75-ohm output impedance. All of the remaining equipment in the receiving terminal is interconnected at this impedance. Beyond the input amplifier the receiving terminal consists of one or more fixed cable equalizers with associated amplifiers, a group of variable equalizers for fine adjustments, and a receiving amplifier which delivers the 1-volt output signal.

A clamper operates in shunt with the 75-ohm path within the receiving terminal as shown in the block diagram. It samples the crest level of each synchronizing pulse of video signal and feeds back at the bridging point a signal which substantially cancels any low-frequency interference or distortion which may have been superimposed during transmission.

The various amplifiers in the system are operated at fixed gains. Small adjustments in the operating level are made by means of an attenuator which is calibrated in 0.5-db (decibel) steps.

The terminals diagrammed in Fig. 3 show unbalanced connections of 75-ohm coaxial type for interconnection with the facilities of broadcasters and industrial users. Optionally these terminals may be arranged for 2-wire balanced connections at 124-ohm impedance. The balanced arrangement is generally used by the telephone company at its own points of interconnection.

A nonrepeatered circuit arranged as shown with flat amplifiers is useful up to about 3.2 miles of cable length. This distance may be increased by substituting other output amplifiers and input amplifiers. The increased range is obtained through an amplifier-to-line network which embodies transformer action over the upper part of the 4.5-mc band. The transformer coupling increases the high-frequency gain so as to overcome losses of the longer cable pair. In practice, the range is increased to about 3.8 miles by substituting the output amplifier with one of the network type. For a still longer cable, the input amplifier also is changed to one of network type.

If the circuit is more than 4¹/₄ miles long, one or more repeaters are inserted to keep the individual cable sections within the limits for a single link. A repeater contains elements found in the terminals, namely, an input amplifier, one or more equalizers with intermediate amplifiers, and an output amplifier which delivers the signal to the succeeding link of cable.

It is seldom possible to lay out a circuit with ideal repeater spacings because of the physical conditions encountered, such as the location of buildings. A practical advantage of *A2A* is its adaptability to irregular repeater spacings and to various cable lengths. The need for this is evident in Fig. 4 which shows examples of actual lines in New York City. *A* and *B*

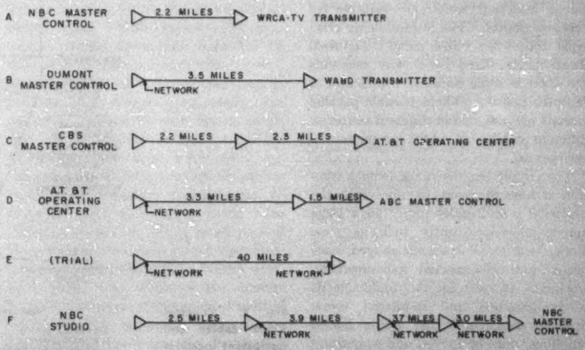

A
NBC MASTER CONTROL
2.2 MILES
WRCA-TV TRANSMITTER
B
DUMONT MASTER CONTROL
3.5 MILES
NETWORK
WABD TRANSMITTER
C
CBS MASTER CONTROL
2.2 MILES
2.3 MILES
AT&T OPERATING CENTER
D
AT&T OPERATING CENTER
3.3 MILES
NETWORK
1.5 MILES
ABC MASTER CONTROL
E
(TRIAL)
40 MILES
NETWORK
NETWORK
F
NBC STUDIO
2.5 MILES
3.9 MILES
NETWORK
3.7 MILES
NETWORK
3.0 MILES
NETWORK
NBC MASTER CONTROL

Fig. 4. Typical A2A lines

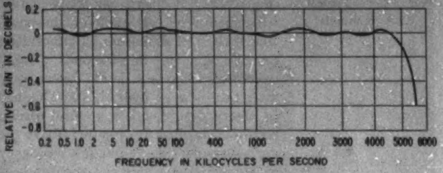


Fig. 5. Over-all gain characteristic of typical 5-mile circuit

are single-link circuits. Line *A* uses the flat amplifiers which have been described. Line *B*, due to its greater length, requires one network-type amplifier at the sending end. Examples *C* and *D* are lines requiring one repeater, and of identical length. The unequal repeater spacing in *D* makes necessary a network-type amplifier for the first cable section. *E* illustrates a 4-mile link which may be operated without a repeater provided two network-type amplifiers are employed, one at each end of the cable. All of these lines have been field-tested successfully and the first four are in commercial service.

An exceptionally long *A2A* line, example *F*, is one of the circuits from a color studio in Brooklyn, N. Y. Theoretically, this line could be designed with only two repeaters instead of three, but here again ideal repeater spacings are not possible. Two of the cable sections are long enough to warrant use of network-type input amplifiers but the equipment shown was selected for the sake of simplicity. There are lines operating in each direction along this route and similar repeaters are employed in all of the circuits.

Performance

The original field trial circuit, 5 miles long with one repeater, was tested over a period of 10 months and confirmed that the design objectives for the system had been realized. After the first commercial circuits were installed on network connections to the New York Television Operating Center and on a studio-transmitter line, their performance was watched critically by the broadcasters and by the telephone company. Aside from a few minor difficulties at first which were traced to faulty components, the performance of these lines in regular service has been most satisfactory, confirming the results of the earlier tests.

The measured over-all gain-versus-frequency characteristic for a representative 5-mile circuit is given in Fig. 5. The residual gain ripples are less than ±0.04 db up to about 4.8 mc. Above this frequency the response rolls off gradually with the 6-db point at about 7 mc. These data were taken with laboratory-type test equipment capable of indicating a difference of $^1/_{100}$ db.

The low-frequency response is plotted down to only 300 cycles. If this curve were extended it would show uniform response down to frequencies below 60 cycles. In the *A2A* system there are no line transformers which could introduce low-frequency distortion.

As mentioned previously, fixed cable equalizers are available in discrete sizes to handle circuits of a variety of lengths. In equalizing a line, values of fixed equalizers are chosen so that the gain at 4.5 mc is brought within ±1.25 db of the low-frequency gain. Precise equalization is accomplished by manually adjusting a group of variable equalizers at the receiving terminal. In this operation, gain is measured relative to that at a reference frequency of 300 kc; this is done at nine specified frequencies while the appropriate equalizers are adjusted. The residual slope across the band is first corrected by setting a cable shape equalizer so that the gain at 4.5 mc equals that at the low reference frequency. Next, the very-low-frequency region is adjusted. Finally, minor deviations within the band are corrected by means of a number of variable equalizers of proper shapes. Altogether, by this procedure ten points in the video band are adjusted to precisely the same relative gain. If the cable sections are long, four additional adjustments are provided making 14 frequency points which may be set precisely. With this equalization method an initial cable distortion of 100 db or more may be reduced to less than 0.05 db. The range of adjustment in the variable equalizers is sufficient to correct deviations on lines up to $6^1/_2$ miles long (120 db of cable slope), with margin for periodic readjustments.

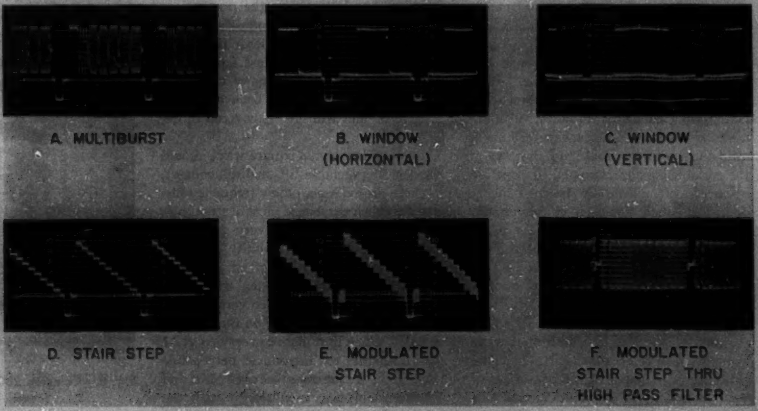


Fig. 6. Test signals over A2A, 9 circuit-miles

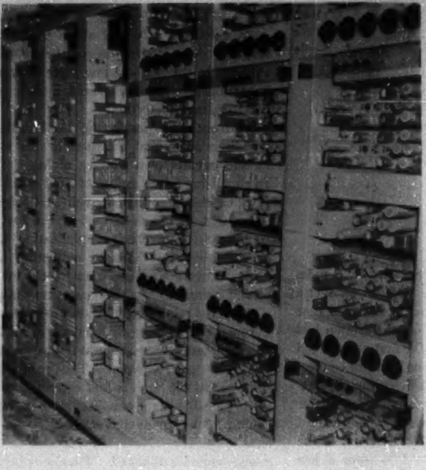


Fig. 7. Typical A2A terminals

On extremely long lines, a second set of variable equalizers is usually added. Such is the case for the previously mentioned 13-mile line which involved an overall cable loss of 238 db at 4.5 mc, corrected by equalization.

The system has proved to be stable in frequency-gain characteristic, the only significant change over a period of time being attributed to the effect of temperature on attenuation of the video cable. With most of the cable underground, this change is a gradual seasonal effect which lags a month or more behind the outside temperature. Based on measurements made over an extended period, the average rate of slope change is approximately 0.15 db per week for a 5-mile line.

Each fixed cable loss equalizer has associated with it a corresponding phase equalizer for correcting the phase distortion in the related length of line. With the appropriate phase equalization associated with each loss equalizer, it is unnecessary to make phase measurements or adjustments in the field. In the case of the variable equalizers no phase correction is necessary. Not only is their phase distortion small but their characteristics above 4.5 mc are such that their phase tends to be complementary to the characteristics being equalized.

Fig. 6 shows test signals photographed after transmission over two $4^{1}/_{2}$-mile circuits in tandem. These are the familiar multiburst, window and stair-step wave forms from a Telechrome portable signal generator.[1] Signals like these are not used for making adjustments in the *A2A* system, but they are convenient for rapidly checking the over-all quality of transmission through the various facilities involved for a network television program.

It is essential, of course, that the oscilloscopes used at various points be alike in performance for wave-form tests to be significant. In this instance, a single high-quality oscilloscope was used, first to set up the signals, and then to observe them over the looped circuits under test. An *A2A* line in normal adjustment will introduce no appreciable change in the multiburst presentation, either as to the relative amplitudes of the white flag and sine-wave bursts or as to their symmetry. A pronounced change in the proportions of the test signal would indicate a probable trouble condition, either in the transmission path or in the test equipment. The effect of various forms of distortion on these test signals has been described elsewhere[2] and will not be covered in detail.

The window-signal square wave should show no appreciable tilt or displacement, as these are evidence of low-frequency distortion which can cause picture streaking. The *A2A* systems shows very good low-frequency response and stability.

High-frequency ringing, if present, would be evident at sudden transitions from black to white, or vice versa, where the transients would show on the window signal. When first placed in service, *A2A* was thought to introduce perceptible ringing at frequencies between 4.0 and 4.5 mc. This was inconsistent with the design theory, and was not supported by subsequent tests. It was found, however, that if the test signal has an abnormally fast rise time (0.1 microsecond), a small overshoot will appear on the square wave after transmission over *A2A*. This will be observed only on a wide-band oscilloscope. The effect is produced above the normal transmission band at a frequency of approximately 5 mc and apparently is due to the phase characteristic in the cutoff region. In the *A2A* system phase equalization is not carried above 4.5 mc inasmuch as the higher frequencies are filtered out by television broadcasting transmitters and thus do not affect home receivers. The spiking effect may be minimized by proper adjustment of rise time in the test signal generator and synchronizing pulse generator.

Differential gain (as checked by the modulated stair-step signal) and differential phase are of special interest because of their effect on the color subcarrier in the National Television Systems Committee color system. The amount of differential gain and phase is, of course, dependent on the number of amplifiers used in each circuit, but does not necessarily add arithmetically when units are placed in tandem. Measurements taken with the *47A* transmission measuring set[3] show

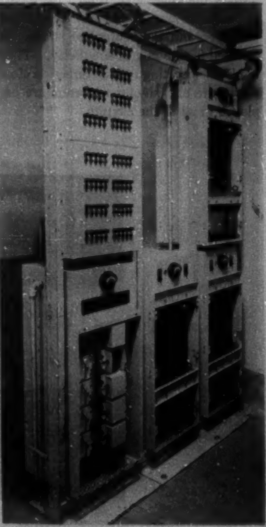


Fig. 8. Common 200-volt d-c power supply (covers removed)

Vol.
Pt. 1
1956

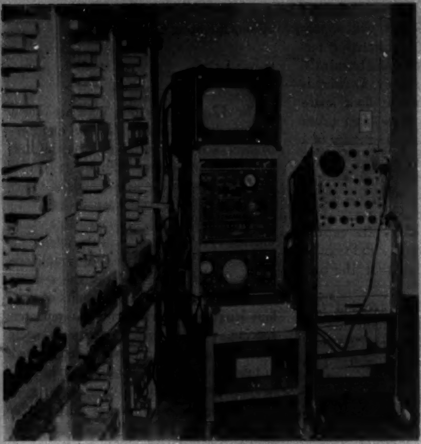


Fig. 9. Typical test equipment

that on a typical *A2A* circuit, differential phase is 0.2 degree or less, and differential gain of about 0.2 db may be expected, for the maximum signal voltage excursion. Measured on a voltage basis, by the modulated stair-step method, the differential gain is less than 5-per-cent white expansion. These values are small fractions of the allowable differential phase and gain, from camera to receiver, in the National Television Systems Committee system.

Signal to Noise

Noise objectives for *A2A* contemplate, as in other aspects of performance, that the local circuit is part of the tandem network connection. By design, the *A2A* line must have a better signal-to-noise margin than is accepted as satisfactory for the over-all network. The effect of random noise on a picture depends on the spectral distribution of the noise across the video band. With appropriate weighting the presently accepted objective for a long network is about 54 db.[4] This is in terms of peak-to-peak signal to weighted rms noise. It includes requirements imposed by color which are somewhat more severe than monochrome. Considering the number and length of local circuits which may be involved, an objective of 65 db was taken as desirable for one local circuit. This is the design figure which determines the allowable length of cable sections in the *A2A* system. Field measurements of weighted random noise are in good agreement with the computed performance.

In field operations, this type of weighted noise measurement is not commonly made. Noise is usually judged by visual observation of picture or by checking the video wave form for thickening of base lines in the blanking interval. Indication of noise by the oscilloscope method depends on the characteristic of the instrument used. For example, a wide-band oscilloscope with flat response exaggerates the effect of noise because it lacks the frequency weighting and also displays noise effects which are outside of the video band.

The *A2A* video system contains numerous safeguards against extraneous noise. Among these are balanced-line transmission, elaborate shielding, and shock mounting of amplifier panels. Wire video transmission, by its very nature, involves the amplification of weak signals at the end of long sections of cable, where the higher frequencies have been attenuated as much as 67 db below 1 volt. Amplifiers handling these low-level signals are susceptible to outside disturbances and require that certain precautions be taken. Noise problems may be avoided by installing the equipment so that it is not subjected to abnormal mechanical vibration, it is not subjected to abnormal acoustical noise, and it is separated a safe distance from possible sources of electrical disturbance. The last consideration includes electric equipment containing make-and-break contacts such as signaling devices, relays, and switches. Normal precautions are also required to bond and ground equipment frameworks and to eliminate differences of potential between the *A2A* ground, the power-line neutral, and the ground used for studio equipment.

Typical Installations

The following will consider typical *A2A* equipment from the point of view of one

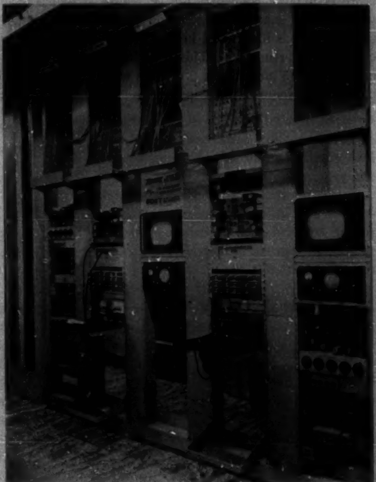


Fig. 10. Jack fields and test positions

planning new installations or having to operate the system. Fig. 7 shows a typical group of terminals as installed at a studio location. The racks are self-supporting, 7-feet high with 19-inch equipment panels. Wiring is enclosed in the vertical rack members which are of duct construction. *A2A* equipment is specially designed for duct-type racks and generally it is not practical to mount the panels on other types of framework. In this illustration the two left-hand bays contain power regulator and alarm panels, and selenium-type rectifier units. The third bay contains six transmitting terminals. At the right are six receiving terminals, two per bay. These facilities should be located within 100 or 200 feet of the broadcaster's jack panel with which they are connected by means of flexible coaxial cables. Room for telephone company equipment may be determined in the early stages of planning for new studios. In established locations, where space is usually at a premium, considerable ingenuity may be called for on the part of the broadcaster and the telephone company to find a suitable location which meets operating requirements.

A2A terminals are also available in portable cabinets. These are convenient when the service is of temporary nature or or where only one or two lines are involved.

The *A2A* system employs small rectifier units which in general supply plate power to an individual receiving terminal, an individual repeater, or up to four transmitting terminals. For locations where there is a large concentration of equipment, use is made of a common power supply such as illustrated in Fig. 8 with cover panels removed. This unit has a capacity of 15 amperes at 200 volts with provision to increase the capacity to 22.5 amperes by adding another rectifier section. A 15-ampere unit will supply power to a total of 22 transmitting terminals and 22 full-size receiving terminals. With a common power supply, a saving is realized in cost and in the floor space occupied. Precautions have been taken in the design to minimize the chance of a service interruption. Should one rectifier unit fail, lose its a-c supply, or require servicing, the remaining units will carry the full load without interruption. In a similar manner sections of the filter capacitor are protected individually and may be replaced during the time that the unit remains in service.

An elaborate array of precision test equipment is provided for circuit line-up, equalization, and maintenance. Some of the typical gear is shown in Fig. 9. This illustrates the need for adequate working space around the terminals. All of the testing preferably is done right at the terminal equipment. Allowance is made for the short coaxial to the studio termination, by substituting an equivalent loop of the same cable during the tests.

Fig. 10 shows a large *A2A* installation in a telephone building indicating the central jack fields and test positions. Here, the high ceilings permit use of 11-foot 6-inch racks. The video cables terminate at the top of the bays where the pairs are connected by flexible patch cords to the assigned equipment. Secondary jacks are placed at working level. Here circuits may be monitored and, in the event of trouble, emergency units may be patched in. The telephone key and lamp panels form part of a communication system which connects with other video locations throughout the city. Some of the test equipment is rack-mounted; the remainder, not shown, is portable. This installation provides terminals for 36 *A2A* circuits serving the New York Television Operating Center.

Conclusions

The *A2A* video system fills a need for a high-quality local wire facility to interconnect television studios, transmitters, and other operating points. It transmits standard, composite signals at video frequencies for distances up to about 10 miles over shielded low-loss cable pairs provided for the purpose in the telephone plant. Unit construction of amplifiers and cable equalizers adapts the system for flexible operation with circuits of various lengths and with irregular repeater spacing. Transmission quality is such that several *A2A* lines may be connected in tandem with intercity facilities in normal network operation.

Once the video cables and *A2A* equipment have been installed, circuit line-up is straightforward and rapid and lines can be rearranged readily to meet changing service demands. Careful planning of *A2A* facilities with particular emphasis on good working space and proper location will facilitate maintenance and assure optimum performance of the system.

References

1. ROUTINE TESTS AND LINEUP PROCEDURES FOR AMPLIFIERS AND CIRCUITS USED IN COLOR TV, H. C. Gronberg. National Broadcasting Company, New York, N. Y., May 1954.

2. TELEVISION SIGNAL ANALYSIS. American Telephone and Telegraph Company, New York, N. Y., May 1955.

3. DIFFERENTIAL PHASE AND GAIN MEASUREMENTS IN COLOR TELEVISION SYSTEMS, H. P. Kelly. *AIEE Transactions*, vol. 73, pt. I, Nov. 1954, pp. 565–69.

4. THE MEASUREMENT OF RANDOM MONOCHROME VIDEO INTERFERENCE, J. M. Barstow, H. N. Christopher. *Ibid.*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 735–41.

Multiunit Neutralizing Transformers

T. F. PALMQUIST
ASSOCIATE MEMBER AIEE

IN RECENT YEARS there has been a large increase in the number of communication, telemetering, pilot-wire relay, and other control circuits required by power utilities in Canada. Service continuity on these circuits is of prime importance, and in many cases neutralizing transformers are required to maintain this continuity and to prevent damage to the circuits or terminal apparatus due to ground potential rise.

Ground potential rise occurs whenever a power-system fault, or lightning, causes current to flow through a power-station grounding structure to earth, the amount of rise being dependent upon the current flow and upon the resistance to ground of the power-station grounding structure. The amount of current flowing to earth through the station grounding structure depends, of course, on the location of the fault in the power system and the type of power system including the station grounding arrangements, transformer impedance and connections, and proximity of other grounded structures. In commonly occurring situations, the resistance to ground of a power-station grounding structure may vary from small fractions of an ohm to several ohms and fault currents impressed on a station ground may vary from a few hundred to many thousands of amperes. In some circumstances, the maximum fault current may be controlled by the insertion of additional impedance in the fault-current path, such as the provision of reactors in the ground connection to wye-connected transformers, thus limiting the ground potential rise.

Increases in the size of generating, transformer, and other power stations has resulted in consequent increases in potentials, which may be applied to communication and control circuits serving a power station at times of fault on the power system. This has necessitated development of neutralizing transformers with increased neutralizing capacity. These neutralizing transformers are usually inserted in the communication or control circuits at the power station concerned and their action, as described further in this paper, prevents excessive currents and voltages from being impressed on the circuits, thus preventing protector operation and possible circuit failure.

In Canada there have been projected circuit requirements of 100 or more circuits terminating in the larger generating stations. Stations with such large circuit requirements are usually major hydro developments, some of which are associated with the St. Lawrence Seaway project.

Previously, multiple installations of single-unit neutralizing transformers were used to provide protection. This, however, increased mounting-space requirements and complicated the wiring arrangements with consequent increases in costs and hazards. Multiunit neutralizing transformers are quite often a reasonable solution to problems of this type as the space requirements are much reduced and the wiring is simplified. Their possible range of rated neutralized voltage is adequate for protection against potential differences which would otherwise exist between the communication or control circuits and ground. Designs for multiunit neutralizing transformers with rated voltages exceeded about 25,000 volts are not as yet available.

Operation of a Neutralizing Transformer

The operation of a neutralizing transformer, in preventing grounding of a communication circuit through its protectors when potentials are applied, is briefly described here for sake of completeness. Fig. 1 illustrates a typical communication circuit serving a power station with a neutralizing transformer in the circuit. A power-circuit fault to ground will cause current to flow through the power-station grounding structure and raise the potential of this ground with respect to grounds at remote points. This potential will be applied to the communication circuit between the power-station protector ground and the central-office protector ground and without supplementary protection may cause these protectors to operate and perhaps permanently ground or damage the communication circuit. With the application of a neutralizing transformer as shown in Fig. 1, the power-station ground-potential rise is applied to the primary winding of this transformer and the resulting current flow induces an equal and opposite potential in the secondary windings. As the secondary windings are connected in series with the communication circuit, this induced voltage will reduce the potential applied across the protector blocks to zero under theoretically ideal conditions and, practically, to less than the operating rating of these protector blocks. Essentially, the neutralizing transformer applies a potential across the circuit protectors which is equal and opposite to that applied to the protectors by ground potential rise.

Description of Typical Multiunit Neutralizing Transformer

Multiunit neutralizing transformers may be obtained in any size from about two to about 101 pairs. The primary and secondary windings of the transformer described in this paper are wound parallel on the transformer core, the primary winding consisting of no. 8 American Wire Gauge polyvinyl chloride (PVC) insulated wire and the secondary consisting of a 26-pair inside-wire cable with 22-gauge conductors insulated with PVC, a core wrapping of two layers of Mylar tape each 2 mils in thickness, and with a PVC outer jacket. Transformers of sizes larger than 101 pairs are somewhat impractical as the cable becomes quite large and difficult to wind and there is increased leakage reactance due to generally increased primary and secondary winding separations. As the designs are similar, a 26-pair transformer has been selected as typical and the descriptions and test data included in this paper will refer to this size of transformer. The voltage rating of this transformer is 5,000 volts at 60 cycles.

Figs. 2 and 3 show the 26-pair transformer; in Fig. 2 its case is removed. This transformer weighs approximately 4,400 pounds. Figs. 4–6 show three views of the transformer mounted on channel beams supported by two poles. As the transformer in this installation was not grounded at its location, the transformer base was placed about 8 feet above the ground. This was done to prevent the

---

Paper 56-929, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted July 16, 1956; made available for printing July 19, 1956.

T. F. PALMQUIST is with the Bell Telephone Company of Canada, Montreal, Que., Canada.

The author acknowledges the assistance of J. B. Hays and L. H. Sessler of Bell Telephone Laboratories, E. J. Bonnessen and A. F. Minor of American Telephone and Telegraph Company and G. Y. R. Allen of Osborne Electric Company, Toronto, Ont., Canada, in the preparation and review of this paper.

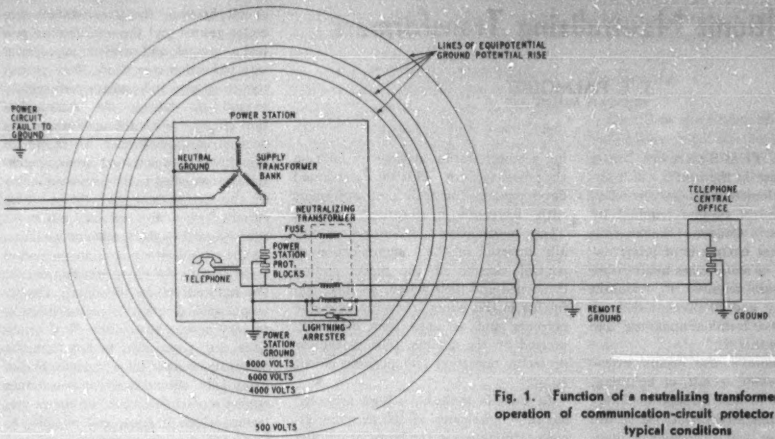

LINES OF EQUIPOTENTIAL GROUND POTENTIAL RISE
POWER CIRCUIT FAULT TO GROUND
POWER STATION
NEUTRAL GROUND
SUPPLY TRANSFORMER BANK
NEUTRALIZING TRANSFORMER
FUSE
POWER STATION PROT. BLOCKS
TELEPHONE
LIGHTNING ARRESTER
POWER STATION GROUND
8000 VOLTS
6000 VOLTS
4000 VOLTS
500 VOLTS
TELEPHONE CENTRAL OFFICE
REMOTE GROUND
GROUND

Fig. 1. Function of a neutralizing transformer in preventing operation of communication-circuit protector blocks under typical conditions

installation being touched from the ground as this would be hazardous at times of fault in the power system. In most installations the neutralizing transformer is grounded at the location at which it is installed, thus removing this hazard and the necessity for special protective measures.

Figs. 5 and 6 illustrate the complexity of wiring and the large mounting-space requirements where a number of single-unit neutralizing transformers are placed at one location. Protective coverings may be required for the neutralizing transformers where they are connected in series if there will be differences in potential between transformer cases or between transformer cases and ground. These coverings may also be required for frame terminations should there be large differences in potential between the various terminations concerned, such as where all connections to a neutralizing transformer terminate on a frame. Special care must be taken in such instances to provide adequate dielectric strength between termination mountings and to provide means of short-circuiting the transformer primary winding to eliminate potential differences between the transformer terminations when the cover is removed for maintenance or other purposes. In the illustrations the covers were made from plexiglass. Complex installations of this type are eliminated by the use of multiunit neutralizing transformers.

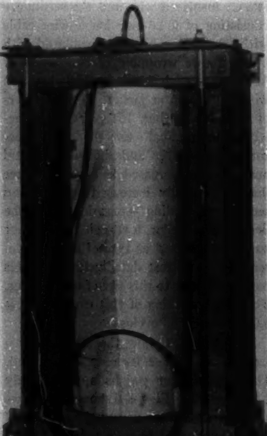


Fig. 2. 26-pair 5,000-volt neutralizing transformer with case removed

Types of Circuits as They Affect Neutralizing Transformer Design

Of the various types of circuits normally in use at the power station, such as telemetering, pilot-wire relay, load control, and voice, only those employing grounded direct current will effect neutralizing transformer design. All of these services may employ grounded direct current; however, the continuous current in any conductor of a leased facility is limited to 0.35 ampere. The longitudinal direct currents will increase the flux density and tend to saturate the transformer, depending on the a-c excitation of the core. This will decrease the impedance of the primary winding and thus increase the remnant voltage as additional current will flow through the primary circuit. An increase in the size of the transformer core, alterations

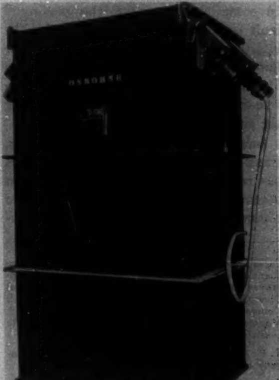


Fig. 3. 26-pair 5,000-volt neutralizing transformer weighing slightly over 4,000 pounds

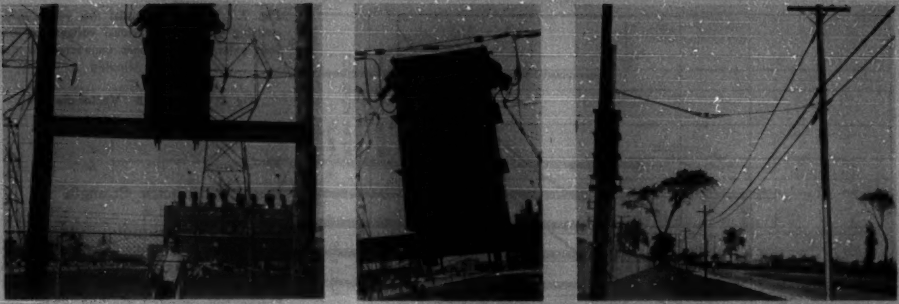


Fig. 4 (above across). Multiunit neutralizing transformer installation

to the type of steel core, or variations in wire size and number of turns may often be introduced advantageously in transformer design to minimize this problem.

Primary Winding External Impedance and Remnant Voltage

The exciting current in the primary winding of the neutralizing transformers must flow through the connection to the remote ground. The remnant or unneutralized voltage is the vector sum of the voltage drops across the remote ground resistance, the connecting conductor between the power-station ground and the remote ground, the primary resistance, and the leakage reactance of the transformer. The primary resistance and leakage reactance are, of course, determined by the transformer construction. Thus the resistance of the remote ground-protector lead wire and the resistance to ground of the remote ground electrode must be sufficiently small that the remnant voltage resulting from the voltage drop across these resistances will not be excessive.

The remnant voltage of the 26-pair 5,000-volt multiunit neutralizing transformer with 0- and with 35-ohm external impedance in series with the primary winding, is shown in Fig. 7. A vector diagram of the various voltages and currents involved in determining the remnant voltage is shown in Fig. 8. The remnant voltage is, of course, applied longitudinally to a communication facility.

In the experience of the author's company, the types of terminal equipment generally used will not be adversely affected by a remnant voltage of 150 volts.

Fig. 5. Multiple installation of single-unit neutralizing transformers in tandem. Note plastic transformer covers

Fig. 6. Multiple installation of single-unit neutralizing transformers. Note plastic mainframe cover

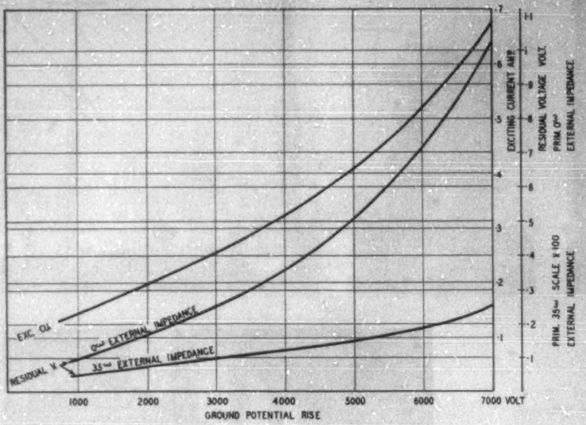
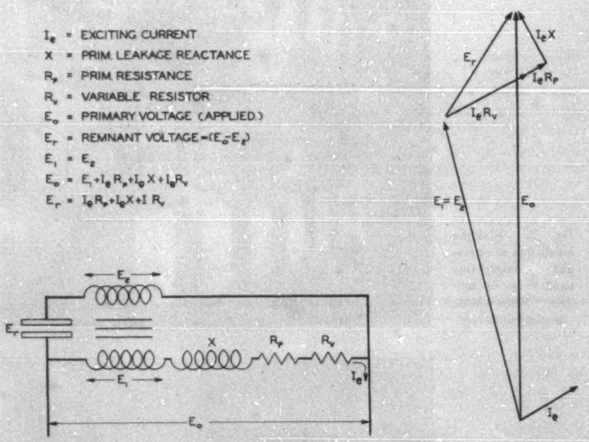


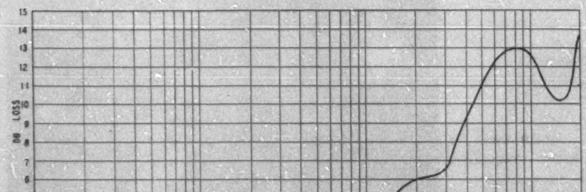
Fig. 7. Ground potential rise versus magnetizing current and residual voltage for various external impedances connected to the primary winding

Fig. 8 (left center). Vector diagram

Fig. 9 (lower left). Insertion loss

I_e = EXCITING CURRENT
X = PRIM. LEAKAGE REACTANCE
R_p = PRIM. RESISTANCE
R_v = VARIABLE RESISTOR
E_o = PRIMARY VOLTAGE (APPLIED)
E_r = REMNANT VOLTAGE $=(E_o-E_2)$
$E_1 = E_2$
$E_o = E_1+I_eR_p+I_eX+I_eR_v$
$E_r = I_eR_p+I_eX+IR_v$





Impulse Dielectric Strength Requirements

To prevent lightning damage to a neutralizing transformer, it is necessary to provide a lightning arrester bridged across the primary winding. The impulse dielectric strength between the windings and between the windings and the transformer core must exceed the impulse rating of the lightning arrester. Distribution-type lightning arresters are normally used with unit-type neutralizing transformers as the use of more expensive arresters is not warranted by the cost of the neutralizing transformer installation. For this reason it is necessary to design unit-type neutralizing transformers with large impulse dielectric strengths. The use of station-type arresters offers economic advantages in the construction of multiunit neutralizing transformers as their impulse ratings are lower than the distribution-type lightning arresters, which permits lower transformer dielectric strengths for a given rms voltage rating and thus economies in construction.

The transformer described in this paper withstood the impulse tests specified in American Standards Association or National Electrical Manufacturers Association standards for a 5,000-volt distribution transformer.

Transmission Losses

The insertion loss of a typical 26-pair 5,000-volt 60-cycle multiunit neutralizing transformer is shown in Fig. 9 and the test arrangement used is shown in Fig. 10. The d-c loop resistance of the secondary windings was 57 ohms for this transformer.

Crosstalk

Typical voice - frequency crosstalk values for a 26-pair 5,000-volt 60-cycle multiunit neutralizing transformer are shown in Fig. 11. The crosstalk was measured using a multifrequency source of disturbing power and a noise measuring set.. The difference in level in decibels between disturbing and disturbed circuits may be determined by subtracting

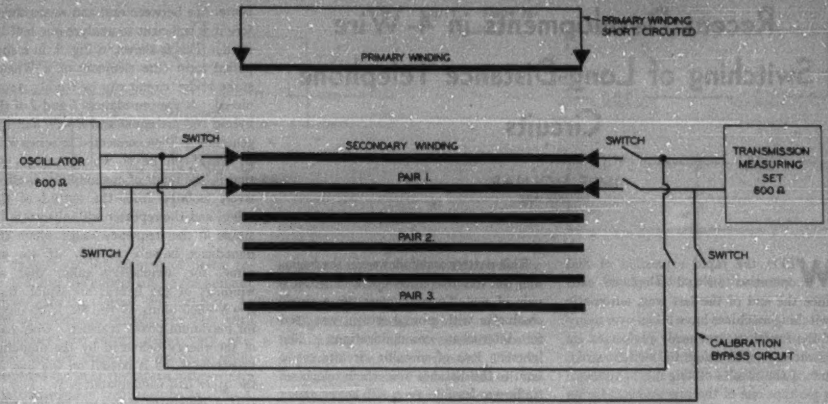

PRIMARY WINDING SHORT CIRCUITED
PRIMARY WINDING
OSCILLATOR 600 Ω
SWITCH
SECONDARY WINDING
PAIR 1.
PAIR 2.
PAIR 3.
SWITCH
TRANSMISSION MEASURING SET 600 Ω
SWITCH
SWITCH
CALIBRATION BYPASS CIRCUIT

Fig. 10. Measurement of insertion loss

the measured crosstalk in decibels from 90 decibels.

The crosstalk results, evaluated in terms of expected crosstalk on standard communication cables, are quite satisfactory.

Leakage Reactance, Winding Resistance, and Capacitance Between Primary and Secondary Windings

The 26-pair 5,000-volt 60-cycle neutralizing transformer selected as typical had the following characteristics:

Leakage inductance = 1.7 millihenrys
Primary resistance = 1.3 ohms
Secondary loop resistance = 57 ohms
Average capacitance between primary and secondary windings = 0.04 microfarad

Experience and Operating Results

There are six existing installations of 26-pair multiunit neutralizing transformers in Canada and a number of future installations are planned. There have been no difficulties in operation or maintenance to date.

References

1. NEUTRALIZING TRANSFORMERS. ENGINEERING REPORT No. 44, Joint Subcommittee on Development and Research. *EEI Publication H12*, Edison Electric Institute, Bell Telephone System, New York, N. Y.

2. PROTECTION OF WIRE COMMUNICATION FACILITIES SERVING POWER STATIONS AND SUBSTATIONS, T. W. Alexander, Jr. *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 587–92.

TYPICAL CROSSTALK VALUES

CROSSTALK VALUES WERE TOO SMALL TO BE MEASURED EXCEPT FOR THE FOLLOWING PAIRS:

PAIRS	DBX
1 - 21	15
2 - 3	23
4 - 8	12
5 - 6	8
5 - 25	12
6 - 21	20
7 - 8	14
9 - 18	18
9 - 25	11
10 - 14	22
12 - 17	8
13 - 27	17
14 - 15	26
17 - 26	15
20 - 23	12
21 - 23	9
22 - 26	11
23 - 25	9

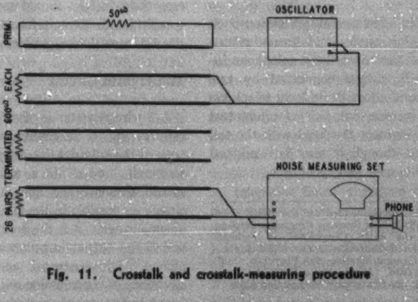

50ω
PRIM.
OSCILLATOR
26 PAIRS TERMINATED 600ω EACH
NOISE MEASURING SET
PHONE

Fig. 11. Crosstalk and crosstalk-measuring procedure

Recent Developments in 4-Wire Switching of Long-Distance Telephone Circuits

IMRE MOLNAR
MEMBER AIEE

WITH the rapid expansion of dial operation on toll telephone calls since the end of the last war, automatic switching machines have taken over many of the functions previously performed by operators on convenient toll switchboards. One of the results of this mechanization, or perhaps one of the prerequisites for its economic justification, was that a varying number of toll circuit links, from one to eight, may be used to build up a connection between two terminal points. It should be understood that the number of links may vary within these limits on different occasions, even between the same two end points and on the same type of calls, at the discretion of the machines and governed automatically by the prevailing traffic conditions. Furthermore, in addition to the maximum of eight links in a toll connection, there will usually be at least one more link at each end, connecting the toll offices with the local offices serving the calling and called subscribers, and there will always be one more link at each end connecting the subscriber's telephone with its local office. This situation differs considerably from that which existed with previously used methods of toll switching; it introduces new transmission problems requiring new engineering solutions.

The Application of 2-Wire Circuits

Except in a few special applications such as power-line carrier and ground-return rural lines, the subscriber's telephone instrument is always connected by two metallic conductors to its local telephone office. Likewise, one pair of conductors is used to connect the local with the toll office, even though it may be equipped with loading coils and amplifiers.

This pattern was followed in the beginning of toll telephone operation, and a pair of metallic conductors, or a single conductor with ground return, was used for interurban communications. The inherent loss of metallic circuits set a limit to the distance over which telephone traffic was feasible, even with heavy-gauge conductors and artificial loading. Spanning of long distances required the use of active voice repeaters.

Until recently all amplifiers of practical importance were unidirectional devices: the amplified speech currents were passed only in one direction. To obtain 2-directional or duplex transmission, two amplifiers, one for each direction, had to be combined into one repeater, and inserted into the 2-wire line at the point where gain was needed. As there was only one pair of input and one pair of output terminals available for the repeater, the input of each amplifier had to be in some way connected to the output of its mate. Due to the gain of the amplifiers, circulating currents established with both amplifiers in series would generate oscillation (or "singing") in the repeater, unless special circuit provisions were made to separate electrically the input of one branch from the output of the other branch. Such a device is the hybrid bridge circuit. While the hybrid circuit is well known to transmission engineers its essential principles should nevertheless be recapitulated, since it forms the core of the 4-wire switching problem.

The Hybrid Circuit

A 2-wire repeater is shown schematically in Fig. 1. Z represents the impedance of the external line as seen from the terminals, and N is a network which should simulate the line impedance as closely as possible in the transmitted frequency range. A is the input impedance and B the output impedance of the amplifiers. The eastward and westward directions of transmission might have different impedance characteristics.

The repeater is symmetrical about the center line between east and west; therefore it is sufficient to analyze one half by itself. This is shown in Fig. 2, in a simplified form (the elements of a Wheatstone bridge circuit can be readily recognized). If the impedances *1* and *2* of the hybrid coil are equal and $Z=N$, and if a generator is then connected in series with B, there will be no voltage across A: that is, the input of one amplifier is effectively isolated from the output of the other, and the repeater will inherently be stable in the frequency range where the impedance balance applies. If we assume, for simplicity's sake, that all windings of the hybrid are equal, that $Z=N=600$ ohms, B will be 1,200 ohms for maximum power transfer. Only half of the energy delivered by the amplifier to windings *3,4* is passed on the line Z; the other half is dissipated in N.

If a generator is connected in series with Z instead of B, energy is received from the line. Assume the impedance of A is one half the impedance of N, 300 ohms. The impedance B reflected into winding *1* is one-fourth of B or 300 ohms, due to the 2-to-1 ratio of windings *3* and *4* to *1*. Therefore half of the voltage across the input terminals of the repeater will appear across the input terminals to the amplifier, if there is no current through winding *2* and network N. Indeed, the voltage across winding *2* will be the same as across winding *1*, but in opposite phase, and since the voltage across winding *1* is the same as the voltage across A, the terminals of N are at equal potential, and thus no energy is dissipated in N. Half of the energy delivered from Z to the repeater is passed on the amplifier input A, and half dissipated in the opposing amplifier output B. Again, 300 ohms for A results in maximum power transfer.

Return Loss

Now, obviously, satisfactory stability of the repeater depends on the degree of equality between line impedance and balancing network impedance. If there is any impedance mismatch, current will be reflected from one amplifier output into the input of the mate amplifier. The reflected current will travel back to the distant end of the circuit where the original current started out, and if the propagation time is appreciable, it may result in undesirable echo. If the total gain in the echo path exceeds the total loss of the echo path, the circuit may even become unstable and "sing." The impedance mismatch may be expressed by a quantity called "return loss," which is the ratio,

---

Paper 56-631, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted November 21, 1955; made available for printing April 12, 1956.

IMRE MOLNAR is with the Automatic Electric Company, Chicago, Ill.

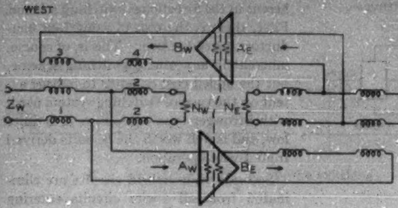


Fig. 1. 2-wire repeater

in decibels, of the amount of power reflected into the hybrid coil to the amount which goes on toward the listener. As line and network impedances vary with frequency, the return losses vary likewise, but for practical purposes a single weighted figure of the return losses within the transmission frequency band is used.

Terminating Impedance

If the repeater is inserted at an intermediate point along a toll line, the networks have to match the toll line impedances only, as seen from the point of insertion of the repeater. While a perfect match is hardly obtainable, the degree of approximation is in a way only a matter of cost. The situation is different with repeaters at the end points of the line, say the "west" terminal in Fig. 1. Here the network N_W should match the impedance of Z_W of the local line. Now, the impedance of this local circuit may vary over a wide range. The circuit is usually composed of three sections in tandem: cabling and equipment impedances in the toll office, the connecting trunk between toll and local office, and the line between the local office and the telephone instrument.

The toll office impedance has little effect in small offices, while in large installations there may be a substantial difference between the different paths a call can take through the switching equipment. The connecting trunk may have practically zero impedance if the local and toll offices are in the same building, or may have 1,500 ohms or more if long and with loading. The telephone line may likewise be very short, and its impedance virtually equal to the impedance of the telephone instrument, or it may be long, and its impedance equal to the characteristic impedance of the telephone line. Moreover, two or more telephones might be connected across the line in a conference, or the line might be entirely open after the subscriber disconnected. With such an extremely wide range and large number of variables, a single value must be assigned to N, which obviously must be a compromise between the differing conditions.

Connecting trunks which are electrically long and have a reasonably stable impedance structure represent a fairly constant impedance, independent of the length of the telephone line terminating them. However, a toll circuit must usually be switched to a variety of connecting trunks, and the network should balance with any of these. Later on, we will see that there are means available to assign individual precision networks for each trunk, or to employ impedance compensation of the connecting trunks. But even then, unless the trunks are long, the effect of the variations in the terminating telephone line prevents a very high terminal return loss. One way to improve the terminal return loss of a connection is to insert a pad between the toll line and the connecting trunk, since the pad will tend to equalize the impedance of the connecting trunk as seen from the toll-line repeater. A 2-db (decibel) pad brings about a 4-db improvement in the terminal net loss, which is usually sufficient in toll offices which have no equipment and cabling irregularities.

Through Switching

When two toll lines are interconnected in a toll office to form a multilink connection, another potential source of impedance mismatch has been created. In such case, referring again to Fig. 1, the west end of one circuit is connected to the east end of the other, and the two compromise networks will be facing each other, with the switching equipment and office cabling interposed. However, as a rule, the match in this case is better than in the terminating condition described in the previous paragraphs.

Because of the undesirable effects of mismatch in the hybrids, the number of repeaters which a 2-wire connection may contain is limited; furthermore, the loss of each section cannot be less than a certain minimum value. The limitations are even more severe on the number and characteristics of intermediate toll switching offices taking part in a built-up connection.

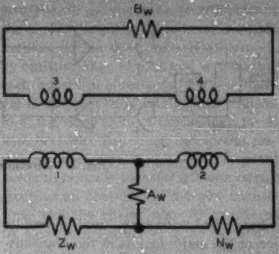


Fig. 2. Schematic half-repeater

Four-Wire Circuits

If the repeater shown in Fig. 1 is cut open between the east and west hybrids, and then the hybrids moved to the east and west terminals of the toll circuit and interconnected by a circuit containing four conductors, with amplifiers inserted along the route where needed, points of return loss are eliminated except at the two end points. This is, of course, the principle of 4-wire transmission, whether over physical conductors or over separate unidirectional carrier channels. However, as long as switching between toll circuits in toll offices is carried out over two wires, both the incoming and the outgoing toll circuits must be reduced through hybrid circuits from four wires to two wires at each intermediate switching office, and the restrictions as to the number and characteristics of such offices in a built-up connection apply as in the foregoing.

The ideal situation, of course, would be to have unidirectional 4-wire circuits in all transmission paths, local and toll, including the telephone line, up to the telephone transmitter and receiver. As a matter of fact, the telephone instrument is a 4-wire device, where the input speech and the output sound are, apart from the circuit arrangement in the telephone, independent of each other. Now, for economic reasons only 2-wire circuits are used for telephone lines; therefore the essentially 4-wire circuit of transmitter and receiver is reduced by an arrangement which is in many respects similar to the one we have discussed for the 2-wire repeater. Telephone instruments have their echo problems just like repeaters; the transmitter current reflected into the receiver

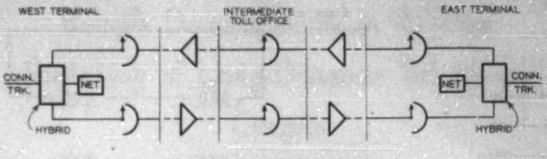


Fig. 3. Directional 4-wire switching

is here called "side tone," and represents one of the major factors in the transmission performance of telephones. The telephone transmitter is a sort of amplifier, but due to the peculiar design and method of use of the telephone handset, there is no acoustical coupling of practical importance between transmitter and receiver so telephones are highly stable in spite of the very imperfect impedance balance of their 4-wire termination. Loudspeaking telephones, on the other hand, with a considerable amount of acoustic coupling between speaker and microphone are unstable unless the return loss in the 2-wire to 4-wire circuit is high. Two factors contribute to the return loss: the precision of impedance match, and the loss in the acoustic path.

Four-Wire Toll Switching

Whether the transmission engineer's millennium of a complete 4-wire system up to and including the subscriber's telephone will ever materialize in commercial telephone systems is still obscure at the present, although it has been done occasionally for special applications. However, as the great majority of long-distance toll traffic is being carried over 4-wire circuits, the majority of built-up connections require interconnection of 4-wire circuits at intermediate switching points. The next logical step is, therefore, the elimination of intermediate 2-wire switching, with its undesirable transmission characteristics, as far as is economically feasible.

As will be recalled, we have derived a 4-wire circuit from a 2-wire repeater, such as shown in Fig. 1, by a vertical cut through the center and displacing the two ends to the extreme west and east terminals, respectively. The two terminals were interconnected by four leads, representing a 4-wire toll circuit. The arrangement is then a 4-wire circuit from end to end, unless in intermediate 2-wire switching offices the four wires have to be compressed to two, and again expanded into four. If, on the other hand, the connection should be carried over four physical conductors through intermediate offices, a true 4-wire circuit would be maintained from end to end, with all points of return loss automatically eliminated, except, of course, at the two terminal points, where the 2-wire connecting trunks are joined.

The Directional Switching Scheme

The directional switching scheme is shown in Fig. 3. To distinguish this arrangement from the network switching method to be described later, it will be referred to hereafter as "directional switching," since a separate pair of leads is used throughout for each direction of speech transmission.

The outstanding features of 4-wire switching are, first, the elimination of impedance mismatch at intermediate points by the elimination of hybrids; second, an added gain of about 6.5 db at each such point, also due to the elimination of two hybrid circuits, which, as we have seen before, consume over 3 db at each coil.

Another feature is the inclusion of a hybrid coil at each 2-wire connecting trunk. Since this coil is always assigned to a certain connecting trunk, the balancing network can precisely match the impedance of the associated trunk line, thus greatly improving the terminal net loss. Though this is an important advantage, it cannot always be fully exploited in practice. If the connecting trunk is short, its impedance depends to some degree on the impedance of the terminating telephone circuit; therefore some compromise still has to be made in the value of the balancing network, which sets a limit on the maximum attainable reduction of terminal net loss. High-loss connecting trunks can be fairly accurately balanced, and some of the available excess gain in the 4-wire terminal can be used to counteract the loss in these trunks; however, the tendency is towards reducing the loss within the connecting trunks themselves. Nevertheless, the possibility of more precision in the balancing of connecting trunks is a definite advantage of 4-wire switching.

There are certain disadvantages inherent in the directional switching scheme. First, there is the cost of using four conductors instead of two. This is, of course, inherent in any 4-wire switching scheme, but the added cost depends to a large extent on the type of switching system used; with Strowger-type equipment it is quite low, and is well worth the benefits derived from 4-wire operation.

Second, while hybrid circuits are eliminated from all 4-wire circuits entering the toll office, a hybrid circuit must be provided for every 2-wire circuit, be it a 2-wire toll line or a connecting trunk or an operator trunk. Admittedly, there may be some large toll switching centers that have to take care of a volume of through switching traffic which is large compared with the volume of traffic terminating locally; in such cases the number of calls switched on a 4-wire basis, not requiring a hybrid, may be larger than the number of calls switched to 2-wire, which do require a hybrid. However, such applications are the exception rather than the rule; usually the local terminating traffic predominates. Therefore, in most cases a greater number of hybrid circuits must be provided with the directional switching scheme than with a scheme where hybrids are associated with 4-wire toll circuits only.

Third, care must be taken that the transmission level in the office cabling and switching equipment shall be about the same in both branches of the 4-wire circuit; otherwise crosstalk might occur. This precludes the use of pads between hybrid and trunk, or hybrid and balancing network—a measure which would otherwise help to improve the terminal return loss; if such pads were provided, the input and output speech levels would be substantially different at the hybrid, and even more so at the switching equipment. With the directional switching scheme all pads are located within the toll-line terminal trunk, away from the switching equipment. Even so, due to the loss through the hybrid circuit, there is a difference of about 6 db between the transmission levels in the two branches, at the switches.

Fourth, inasmuch as we have obtained a true 4-wire circuit with separate paths for each direction of transmission, some special arrangement is needed if we wish to intervene at some intermediate point. For instance, if a switchboard operator with a conventional telephone set arranged for 2-wire connections is needed to be bridged across the 4-wire circuit, a rather complex applique is required. The situation is similar with talking circuits for test personnel.

Vol. 7
Pt. 1
1956

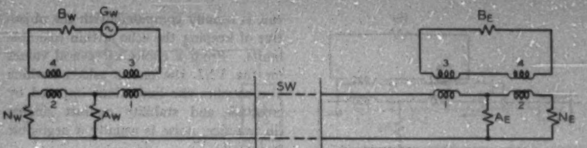


Fig. 4. 2-wire switching

The Network Switching Scheme

The scheme to be described is free from the last three handicaps indicated in the foregoing, but shares all the advantages mentioned, with some more added.

Let us, for a moment, consider 2-wire switching as shown in Fig. 4. G_W is a voice source located in the west terminal. Energy is transmitted towards the east terminal. If the balance is perfect, half the energy from G_W, which is passed through the transformer, is dissipated in N_W, half flows towards A_E (half of the latter is dissipated in B_E) and will be amplified there, but nothing is dissipated in A_W, which means that there will be no echo produced. In practice, however, perfect balancing is impossible due to the switching equipment *SW*, and part of the energy generated at G_W will be reflected back to A_W. For instance, as an extreme case, if one of the leads in *SW* is open, no energy flows to the east, two thirds of the energy passed through the west hybrid is dissipated in N_W and one third in A_W (assuming the same 2-to-1 impedance ratio as before).

As the next step, remove both balancing networks, N_W and N_E, and, instead, connect the terminals as shown in the lower part of Fig. 5. A few facts are immediately apparent. First, assuming that the "line" and "net" paths in *SW* are identical, the bridge on the west side is in perfect balance (if we consider the impedance A_E to have been sliced in two, and one half, together with east winding *2*, moved to the left-hand side of the bridge, Fig. 6). Under such conditions, there will be no potential difference across A_W, and no energy dissipated there. Perfect balance will prevail, as long as any change in the line path is accompanied by an identical change in the net path, and vice versa.

Second, we may now omit A_W altogether, and follow the current flow created by G_W through windings *1* and *2* of the west hybrid coil. If we follow the arrows, we will discover that due to the reversal in the net path, the current converges in opposite phase upon impedance *A*. The east hybrid in this condition is nonreactive, and no energy will be dissipated in A_E. Therefore, all energy delivered by the west hybrid will reach the east amplifier input A_E; this results, in addition to freedom from echo, in a 6-db gain of the circuit as compared with the 2-wire case shown in Fig. 4.

Third, assume B_W to be 1,200 ohms, A_E to be 300 ohms, 1-to-1 transformer ratios throughout, no energy dissipated in transformers and leads, slice A_E into two parts (600 ohms each), and disregard impedances A_W and B_E since they receive no energy. Examining the circuit, we find that generator impedance B_W and load impedance (600 ohms plus 600 ohms) are equal, therefore maximum power transfer conditions are established.

Fourth, in the 4-wire switching scheme being discussed, energy is transferred in both directions over two paralleled 2-wire paths, each path carrying one half of the energy. Therefore, the speech power level will be the same in both paths at any point in the system.

The arrangement just described, employing the network switching principle, is, for obvious reasons, called the "tail-eating" 4-wire switching system. As already explained in the foregoing description, it offers a 6-db gain over 2-wire switching, which is only slightly (less than 1/2 db) lower than attainable with the directional arrangement, due to the inherent transformer losses in the hybrids. Furthermore, it offers the same degree of stability and freedom from echo as the directional scheme, as long as the impedances in the line and net paths are equal. Actually, the circuit is rather insensitive to impedance mismatch; for instance, should one of the four conductors become completely open, there is still about 30-db return loss maintained (versus infinite loss in perfect balance condition).

The tail-eating arrangement is superior to directional switching in that only 4-wire circuits must be equipped with hybrids (which they usually are anyhow), and 2-wire circuits may remain as they are, without adding a hybrid. In contrast to directional switching, the transmission power levels are the same in both pairs of the 4-wire circuit, so there is no crosstalk problem within the quad. Furthermore, by omitting the net path and terminating each hybrid in a network, 2-wire switching is automatically established. Therefore, the same circuits can be used for 2-wire or 4-wire switching; no special provision (other than to open the net path) needs to be made to insert a 2-wire circuit (such as an operator) in a 4-wire connection, and the circuit may be monitored by a direct bridge across either the line or the net pair. The switching circuits of the tail-eating scheme are simpler than those used with directional switching, which simplifies maintenance, and also provides for greater flexibility for possible future changes.

The tail-eating scheme has all the advantages of directional switching. It was mentioned that its inherent gain is about 1/2 db less; however, since there is an excess gain in 4-wire switching anyhow, the pad values are so selected that the transmission losses of both arrangements are the same. This will be explained in more detail later. The tail-eating method also permits the use of precision or compromise networks assigned to individual 2-wire trunks (see Fig. 7), and the presence of pads in both the line and net branches provides for superior terminal return losses.

The only point of caution needed with the tail-eating system is to insure as close symmetry as possible between line and net paths through the switching equipment and central office cabling. In the design of circuits and equipment the symmetry is perfect, limited only by the manufacturing tolerances of the same components. Symmetry in cabling is insured by the use of quadded cables for the 4-wire circuits, and consistent paralleling of the branches in all cases.

Transmission Losses of Toll Circuits

So that we may more fully understand the switching operations employed in a 4-wire switching system, let us divert briefly from our main subject and consider some of the transmission problems involved in long-distance telephone operations.

During recent years the long-distance toll plant has been built up to a point where at present most of the traffic is carried over repeatered voice circuits, carrier channels, and radio links (with the last two media by far predominating) instead of by metallic circuits only, as in the past. It can be safely predicted that the preponderance of derived channels will increase in the future at an even more

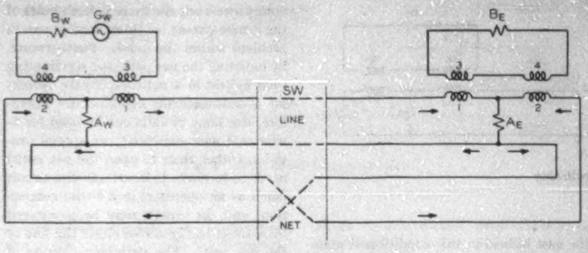


Fig. 5. Tail-eating scheme

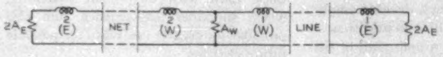


Fig. 6. Tail-eating bridge

rapid rate. Now one of the significant results of this expansion is that, in general, there is more gain available in the toll circuits, in spite of their length, than is necessary to secure the desirable transmission level at their output. The obvious approach in such a situation would be to operate all circuits at zero loss, since, in addition to providing the same level of transmission as obtained by direct interconnection of the terminals, it would make transmission independent of the number of circuits connected in tandem when a circuit is built up between the same two end points over devious alternate routes.

Of course, if the circuits could be operated at a gain instead of zero loss, some of the loss of the connecting trunk between the toll office and the local telephone office, as well as the loss inherent in the line between the local telephone instrument and the local office, could be compensated for. Unfortunately, up to date it has been found impossible to operate long toll circuits on a zero-loss basis. The gain of repeatered voice-frequency circuits is limited by noise and by stability limits at hybrid points, as well as by crosstalk and echo, while carrier circuits must be operated at a finite loss because the last two factors are still appreciable when the distances are long, even though the velocity of propagation approaches the velocity of light.

One of the facts learned from years of operation of transmission circuits is that if two or more circuits are switched together for a through connection, the total loss of the connection can be lower than the sum of the losses of the individual circuits when they are used by themselves. The explanation for this is that the factors mentioned in the preceding paragraph do not increase in proportion to the length of the connection or to the number of circuits connected in tandem. This is particularly true if 4-wire switching is employed at the intermediate points, because then it is only at the terminating points that the 4-wire circuit is reduced to 2-wire; the factor of return loss is eliminated in all but the two terminating offices.

The lowest practicable loss of a toll circuit is the loss required by the foregoing limiting factors when the circuit is switched to another circuit at both ends; it is called "via net loss" (VNL). Experience has shown that on electrically long circuits, echo is the most difficult factor to control, so the permissible VNL of a circuit is usually computed with the objective of keeping the echo within tolerable limits. From a choice of several values for the VNL the one is selected which also satisfies the limitations imposed by crosstalk and stability against singing (in practice, noise is usually a negligible factor).

The computation of the VNL is based on probability considerations, taking into account a large number of contributing elements, such as: length of the circuit and velocity or propagation, the expected precision of balance at both terminals and its statistical distribution, the deviation of the circuit losses from their nominal value and its statistical distribution, the amount of echo tolerated by the telephone users and the statistical distribution of the tolerance, the availability of echo suppressors, and the amount of added loss at the terminating points of the connection. This method of computation can be used only with 4-wire circuits; the VNL of 2-wire circuits is established by engineering judgment based on experience.

Typical VNL values of facilities in use at the present may range from 0.5 db for 50 miles of short-haul carrier to 2.4 db for 200 miles of loaded voice-frequency 4-wire cable, with 500 miles of coaxial carrier about halfway between these two.

The VNL of a built-up connection may be obtained by adding the VNL of the individual sections together, provided that 4-wire switching is used at the intermediate switching points. Intermediate points using 2-wire switching are added potential sources of echo and singing, and it is expected that such connections built up through such points can be worked at the VNL permissible with 4-wire switching intermediate points, only if the average balance (return loss) in each of such offices is 27 db or better. If the balance is lower, the VNL of the circuit terminating there must be correspondingly increased, for instance, by 0.2 db if the balance is only 23 db.

The ultimate objective, of course, is that the VNL of a built-up connection

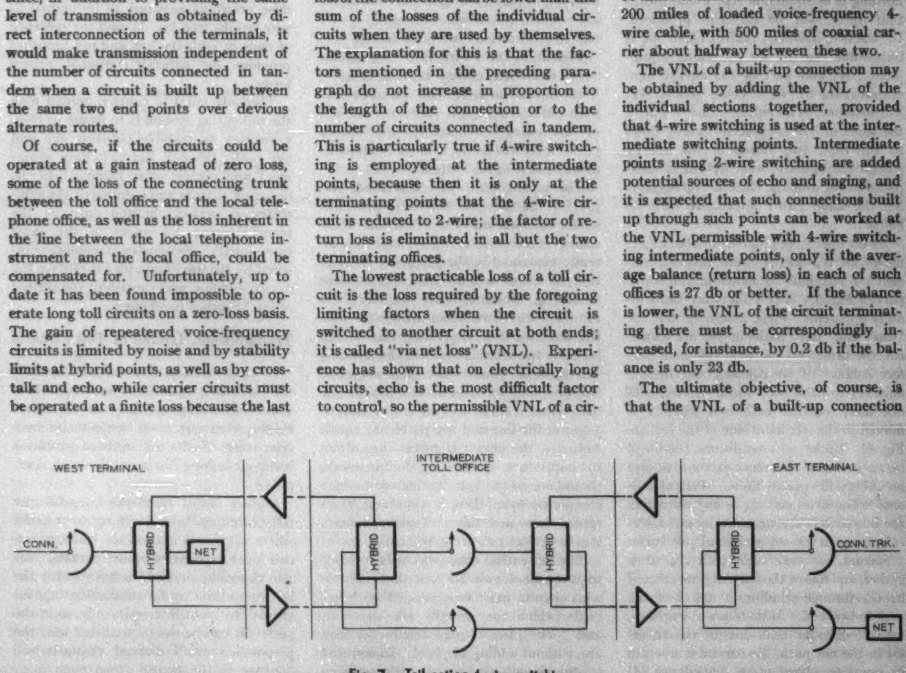


Fig. 7. Tail-eating 4-wire switching

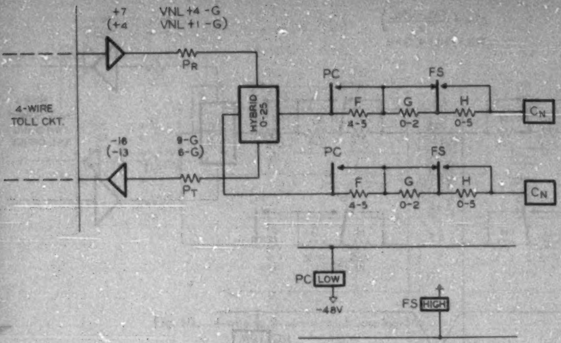


Fig. 8. Idle condition

P_R = receiving pad
P_T = transmitting pad
C_N = compromise net

should be as near as practicable to the VNL of a direct circuit between the same two terminating points. At the present, or even in the foreseeable future, this ideal cannot be achieved. In general, telephone users are likely to object only if the transmission level between the same two telephones varies by more than 5 db on different occasions; therefore the immediate objective is to keep the VNL of a connection within this limit, regardless of which of the possible alternate routes the call may take.

At the present time, we are still quite a way from even this immediate objective. For instance, it may be practicable to maintain circuit stability to such a degree that the net loss of a 4-link connection will be within a range of about 9 db in 96 per cent of the cases, if the variations in losses of the individual circuits are combined on a statistical basis. In an 8-link connection the variation in losses would be about 40 per cent greater. In addition to these, there will be a variation in the average values of the VNL of a connection, dependent on the number and the VNL of the links used for the different alternate routes. Fortunately, only a minor fraction of all calls will use four links in a connection, and the probability of eight links is one in several millions.

Switching of Losses at Terminating Points

In the preceding section we have discussed the via net loss of a connection—that is, the minimum loss which must be maintained due to the inherent characteristics of the links used and regardless of the disposition of the connection at the switching office. The minimum permissible loss is more than VNL when the circuit is used between two telephone instruments, because the same toll circuit must be used under a great variety of telephone loop conditions in the local telephone networks served by the terminating toll offices, and the balancing network in the terminating hybrid has to be a compromise over a wide range of possible impedance values. Therefore, the local line will hardly, if ever, be perfectly balanced, and to maintain echo and stability within satisfactory bounds the over-all circuit loss must be increased. The minimum loss in terminated conditions is called "terminal net loss" (TNL), and at the present state of the art and with the best available facilities TNL = VNL + 4 db, that is, an increment of 2 db at each end.

A toll circuit will, at any given moment, be found in one of the following conditions: 1. It may be idle. 2. It may be switched to a 4-wire toll circuit or to any toll circuit with terminal repeater. 3. It may be switched to a 2-wire toll circuit without terminal repeater or to a local connecting trunk with a high inherent loss. 4. It may be switched to an ordinary local connecting trunk. (In addition, the circuit may be switched through an operator, but this condition will be excluded from the present discussion.) Furthermore, the toll circuit may be a 4-wire circuit (or a circuit having a terminal repeater) or it may be a 2-wire circuit; and it may terminate at the distant end in an office with 4-wire switching or in an office with 2-wire switching. For the present, we will assume that the near end terminates in an office with 4-wire switching facilities. Whether the distant office is of a 4-wire or a 2-wire type may have some bearing on the VNL of the circuit, but not on the pad values used for switching in the near-end office.

In the following we will discuss the ways in which these different states are provided for in the Automatic Electric Company's new *FW-1* Intertoll Switching System. To illustrate the transmission relations, numerical values will be assigned to the various elements. For instance, assume that the transmission level at the output of the demodulator is 7 db above the level of the entrance point of a toll connecting trunk at the other end of the toll circuit link, and that the input level to the modulator is 16 db below the level of the entrance point of a toll connecting trunk in this office. Therefore the total gain of a circuit link from modulator input to demodulator output at the far end is 23 db. These +7- and −16-db levels are accepted standards for modern transmission facilities, though higher total gains (26 db) can occasionally be found, with corresponding changes in the output levels. Older transmission facilities with a total gain of 17 db are still in extensive use with +4- and −13-db levels, respectively. The corresponding values for the transmitting and receiving pads are shown in the figures in parenthesis.

In the discussion assume that the connecting trunk plants are such that TNL = VNL + 4 db is satisfactory. The added loss is derived at each end by a 2-db value of the *G* pad. While at the present time not all connecting trunk plants can assure this value for TNL, an effort is being made to improve the return losses, with this objective. Where the plant cannot be operated at TNL = VNL + 4 db, toll connections may need to have, for the time being, a higher TNL: for instance, VNL + 8 db if the terminal return loss is degraded at each end by 4 db. The 4-db impairment in the terminal return loss can be compensated for by an added loss of 2 db in the transmission path at each end; thus the value of TNL = VNL + 8 db. As the value of the *G* pad in the *FW-1* system is 4 − (TNL − VNL)/2 db, *G* will be 0 db for circuits which must be operated at a TNL = VNL + 8. It is a surprising peculiarity of 4-wire switching that the value of a switching pad is actually decreased when a higher TNL is required and that the additional loss is derived from the trans-

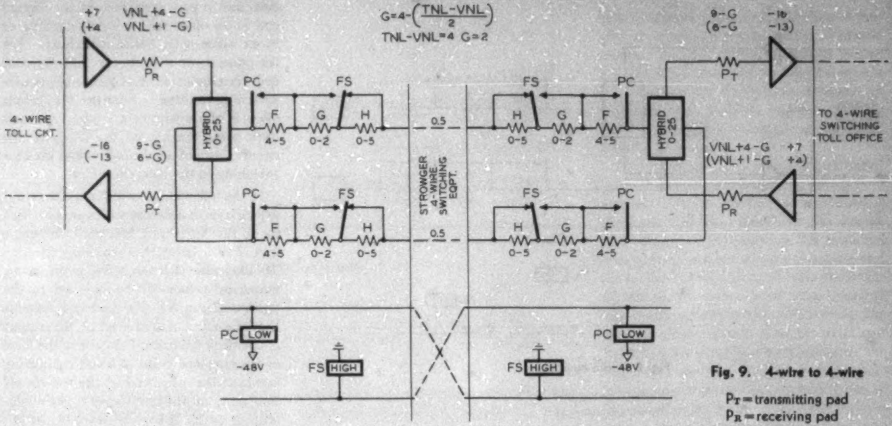


Fig. 9. 4-wire to 4-wire

P_T = transmitting pad
P_R = receiving pad

mitting and receiving pads. The examination of the circuits later in this paper will reveal the reason for this. However, as will also be demonstrated, the terminal return loss of 4-wire switching is inherently high, and therefore, in general, a TNL = VNL + 4 db may be assumed for most applications.

We will further assume that there is only about 0.25-db loss in a hybrid coil in the 4-wire to 4-wire state due to the losses in the associated wiring and the usual transformer losses; the loss is about 3.5 db in the 4-wire state because, in addition to the 0.25 db mentioned, half the input energy is dissipated in the balancing network. Losses in the central-office cabling and the switching equipment may amount to 0.5 db.

1. IDLE STATE (FIG. 8)

When a 4-wire circuit (or a 2-wire circuit having terminal repeaters in the toll office) is not in use, the line and network terminals of the hybrid circuit are terminated in identical networks, and theoretically an infinite terminal return loss is obtained. The 5-db loss inserted between each of these terminals and the networks further improves the balance of the hybrid. When the circuit is seized from the distant end and the connection is built up in the office, the terminating network of the line side is passed on at each step until the terminal of another toll circuit (which will be discussed in the next two cases) or a local telephone is reached, and a very high return loss is continually maintained. When the called telephone answers, the terminating network at the hybrid line terminals is replaced by the impedance of the connecting trunk plus local line terminated by the telephone instrument. The switching details will be described later.

Two-wire toll lines without terminal repeaters have no hybrid circuit and therefore create no problem. They are, however, terminated in the same type of network for the sake of standardization and to improve the stability of any repeater which might be located near the terminal.

2. CONNECTION TO 4-WIRE CIRCUITS (FIG. 9)

When a 4-wire circuit is switched through to another 4-wire circuit, the line terminals of both circuits and the network terminals of both circuits are interconnected as previously explained. The speech energy divides evenly over both paths and then recombines in the correct phase through the hybrid transformers. The only loss inherent in this arrangement is due to the office cabling and equipment; in addition, of course, to the ordinary transformer losses in the hybrids, to a nominal 1-db total, as already pointed out. As the output of a repeater demodulator is at a +7-db level and the input to a repeater modulator should be at a −16-db level, a minimum of 23-db loss must be inserted into both paths. Part of this loss is assigned to the receiving pad, part to the transmitting pad in the repeaters, and the balance (less 1 db) to the switching pads of both trunk circuits.

One of the interesting characteristics of 4-wire switching, as distinguished from 2-wire switching, is that in a through-switch condition, it requires an added loss, while with 2-wire switching in this condition, loss is eliminated as far as possible. This difference occurs because with 2-wire switching half of the available energy is dissipated in the balancing networks, while in 4-wire switching practically all available energy is transmitted from one circuit to the other, thus gaining twice 3 db. Now, assuming that TNL = VNL + 2, with 2-wire switching twice 2 db must be removed from the circuit in the through switch state, while with 4-wire switching (with an excess gain of twice 3 db) twice 1-db loss must be added to switch through with the same VNL as in the 2-wire case. In local terminating condition the TNL-VNL in a 2-wire office is supplied by a 2-db pad or a minimum 2-db trunk loss, while in a 4-wire office it is supplied by the hybrid coil. Since the hybrid loss is 1 db higher than 2 db, the output level of a repeater in a 4-wire office must be 1 db higher than in a 2-wire office, and this difference is then eliminated by the switching pads in a through switch condition.

When two 4-wire circuits are connected together, both *FS* relays operate over the d-c paths, while the *PC* relays remain unoperated due to the high series resistance. The function of the *FS* relay is twofold: it extends the four conductors through the automatic switching equipment, and it replaces the *H* pads by the *G* pads, thus increasing the loss by 1.5 db.

The statement made earlier that the input level to the modulator should be at a −16-db level must be now slightly

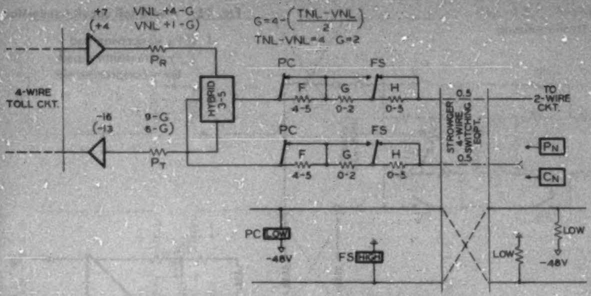


Fig. 10. **4-wire to 2-wire or high loss local**

P_R=receiving pad
P_T=transmitting pad
C_N=compromise net
P_N=precision net

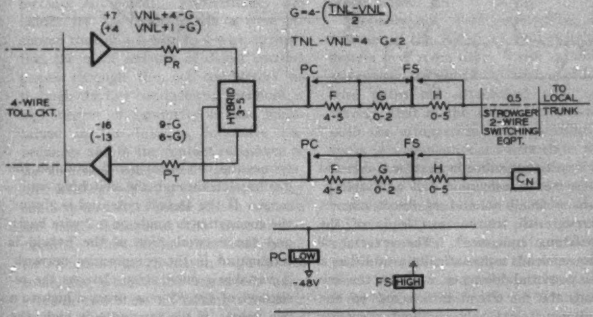


Fig. 11. **4-wire to standard local**

P_R=receiving pad
P_T=transmitting pad
C_N=compromise net

qualified. In a through-switch condition the input level to the modulator should actually be $-(16+\text{VNL})$ db, and the loss in the through-switch path should be $23+\text{VNL}$ db. The values, for the receiving pad P_R and the transmitting pad P_T are so chosen that, together with the switching pads, this amount of loss is obtained, and that the relation $\text{TNL}=\text{VNL}+2$, can also be maintained, as will be shown under case 4.

3. Connection to 2-Wire Circuits (Fig. 10)

When a 4-wire circuit is to be connected to a 2-wire circuit without terminal repeater, it is desirable to improve transmission by utilizing some of the surplus gain available in the 4-wire circuit. Having no terminal repeater, there is no excess gain available in the 2-wire circuit, so there is no purpose in providing pads and switching relays in the 2-wire terminal. The extra gain is derived in the 4-wire terminal instead by cutting out pads *F* and *H*, and inserting pad *G*. Both relays *FS* and *PC* are operated through resistors in the 2-wire terminal. The loss in the connection up to the 2-wire line is $\text{VNL}+1$ db. The additional nominal 1 db is caused by the loss in the hybrid coil, exchange cabling, and switching equipment. It would be possible to compensate for these and to switch the connection with VNL only; however, it is not considered worth while to introduce an additional condition in the pad switching, and special measures to increase the return loss of the hybrid circuit, for such a nominal gain. The balancing network for the hybrid circuit in the 4-wire terminal is supplied from the 2-wire terminal over the net conductors. As a rule, having perfect symmetry between line and network leads, a standard compromise network is sufficient to provide adequate return loss in the hybrid circuit. Should in some cases the structure of the 2-wire circuit cause difficulties, the circuit as shown is arranged for a precision network to balance out the 2-wire line.

The same arrangement may also be used to compensate to some extent for high losses in connecting trunks, by transferring excess gain from the 4-wire terminal. From the transmission engineer's point of view, there is no difference between a 2-wire voice-frequency toll circuit and a high-loss toll-connecting trunk. While the present objective is to keep connecting-trunk losses below 5–6 db, plans are being made to design such trunks in the future on a $\text{VNL}+2$ db basis, or even for VNL where an intermediate switching point in the connecting trunk is involved.

As ordinary 2-wire circuits without terminal repeaters are not equipped with hybrids or pads, they are directly interconnected by the switching equipment, and there is nothing special to describe or to illustrate. Since there is no gain available anywhere, the loss of the through-switch connection will be the sum of the TNL's of both individual links.

4. Connection to a Local Trunk (Fig. 11)

Terminating a 4-wire circuit locally through an *FW-1* 4-wire toll office is fundamentally the same as terminating it through a 2-wire toll office, because the network leads are not extended unless a high-loss connecting trunk is used as was the case in item 3. In a 2-wire office, $\text{TNL}=\text{VNL}+4$ db, and the extra 4-db loss is supplied either by 2-db switching pads at each end of the connection, or by 2-db fixed pads in the connecting trunks, if they have less than 2-db loss. The same relation $(\text{TNL}=\text{VNL}+4)$ applies to a 4-wire office. Because the network leads are not extended, there is no control potential available to operate relays *FS* and *PC*. Pads *F* and *H* provide the loss required to obtain the required TNL.

The local connecting trunk is balanced by a compromise network, incorporated into the 4-wire trunk terminal. Because of the facilities available in the 4-wire switching system, there is altogether 10-db loss between the compromise network and the network terminals of the hybrid,

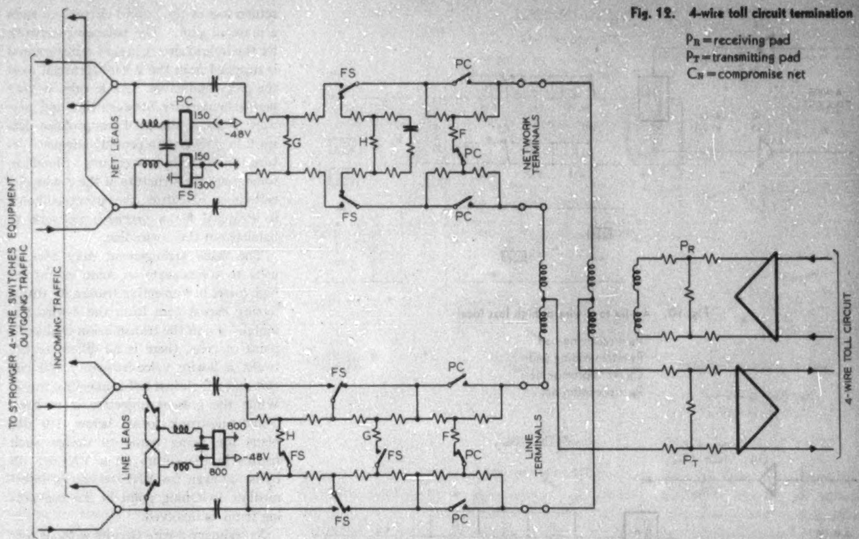


Fig. 12. 4-wire toll circuit termination

P_R = receiving pad
P_T = transmitting pad
C_N = compromise net

and between the line terminals of the hybrid and the connecting trunk. Through such an arrangement a very high terminal-return loss is obtainable which is superior to the conditions prevailing in conventional 2-wire-switching offices where only a 2-db pad is inserted between the hybrid and the connecting trunk.

An ordinary 2-wire circuit is directly connected to a local trunk and there is likewise nothing special to discuss. The TNL of such a connection will simply be the inherent loss of the 2-wire toll circuit.

Four-Wire Circuit Termination (Fig. 12)

In preceding sections, the transmission principles of the *FW-1* intertoll switching system with 4-wire switching were described. Some of the pertinent operations will be explained in the following.

All such operations are carried out over the two net leads, after they are extended through the automatic switching equipment. If a connection is to be extended on a 4-wire basis, line pairs and the net pairs of the two circuits are tied together directly without any attachments. As may be seen from Fig. 12, the line leads are interconnected straight, while there is a reversal between the net leads, to obtain the correct phase relation in the hybrid circuit.

Experience indicates that for stable and noise-free transmission it is advisable to pass a small amount of direct current through all transmission leads of the switching equipment. The reversal of the net leads automatically establishes a d-c potential difference, while for the line leads the d-c circuit is reversed on one (outgoing) end only, without reversing the voice transmission path. The d-c portion of the line leads is used for various switching functions not pertaining to the present subject, and at the same time maintains about 30 milliamperes of current in each lead at all times.

The conditions under which relays *FS* and *PC* are to operate were discussed previously. Relay *FS* will always operate when a connection is to be made through the switching equipment on a 4-wire basis. In its unoperated position, *FS* terminates the network side of the hybrid in a compromise network, while in the operated position it extends the network terminals through the switching equipment. When operated, *FS* replaces the *H* pads by the *G* pads. The only function of the *PC* relay is to cut out the *F* pads in its operated position.

All circuits which may provide 4-wire transmission, such as cases 2 and 3, must have negative 48-volt potential on the winding of the *PC* relay. Under such condition the *FS* relays at both ends of the connection can operate and establish the 4-wire path through the switching equipment. If the 48-volt potential is absent, the connection is made on a 2-wire basis, and the network side of the hybrid is terminated in the compromise network. A variable ground strap changes the resistance of the *FS* relay from a high to a low value; if the resistance is high, the *PC* relay at the opposite end of the connection cannot operate. This is the condition on a 4-wire to 4-wire connection, described in case 2. The *FS* relay must have a low resistance if it is used in a 2-wire circuit that includes the balancing network for the opposite terminal, as described in case 3. The low resistance of *FS* will then permit operation of relay *PC* in the opposite 4-wire terminal. (As already mentioned, in 2-wire circuits with the *FS* and *PC* relays, ordinary resistors may be substituted to control the relays in the opposite terminal.)

It will be noted from this circuit that the line and net conductors are perfectly symmetrical, and since this symmetry is inherent in the central-office cabling and in the physical and electrical layout of the switching equipment, no special measures need be taken in the central-office equipment to obtain a high terminal-return loss even when the connection is made on a 2-wire basis. Quadded switchboard cables are used for the line and net

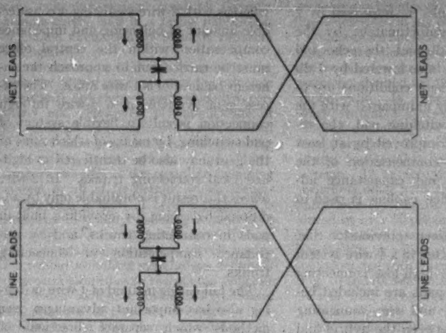


Fig. 13. A-c-d-c phasing circuit

conductors, and these conductors consistently follow identical paths, even where tie connections are made to intermediate equipment. Furthermore, the values for the control relays were so chosen that in a 4-wire to 4-wire connection the same amount of direct current passes through the net conductors as through the line conductors; therefore, a high degree of symmetry is also maintained in the bridging inductive impedances and in the contact resistance of the switching equipment. To keep contact resistances low, precious-metal relay contacts and silver-inlaid switch contacts are used throughout the transmission circuits. When the low-resistance option for *FS* is used, the current through the net conductor is somewhat increased; however, in this state no talking current is passing through the net circuit, and minute variations of the bridging impedance in the net circuit (due to a change in d-c saturation) are inconsequential.

When a call between two 4-wire circuits is originated by an operator within the office, both of these circuits are reached through their terminals assigned for outgoing traffic. In such a connection an additional reversal of the net leads is required, as well as a reversal of the direct current without simultaneously reversing voice transmission in the line leads. The circuit employed for this purpose is shown in Fig. 13. There are two separate transformers for the two sets of leads, connected in such a way as to maintain d-c continuity. The poling of the transformer in the net circuit is such that there is no reversal in phase of the voice transmission. The physical reversal of the conductors creates, then, the required reversal of both d-c and voice-current paths. The poling of the transformer in the line circuit is such that there is a reversal in phase of the voice transmission. The physical reversal of the conductors, then, creates the required reversal of the d-c path and restores the voice current to its proper phase.

Probably the most outstanding advantage of the *FW-1* intertoll switching system over more usual 4-wire switching systems is the simplicity with which 4-wire paths are established and, conversely, that in the absence of a 4-wire path the arrangement automatically returns to a conventional 2-wire state, with better than normal return-loss characteristics. As has been shown, the 4-wire state is contingent on the operation of the *FS* relay, which can take place only if a control potential is extended from the opposite terminal over one of the net conductors. Therefore, a 2-wire state is retained if the opposite terminal is not arranged to extend the control potential; or if the intermediate switching equipment is not arranged to extend the net leads; or if the connecting path is open due to some trouble condition. In other words, the 2-wire state exists in all cases unless the conditions for the 4-wire state are positively established (for instance, if an operator or test personnel intervene in a 4-wire connection, the net leads are simply disrupted and the connection automatically reduces itself to a 2-wire state, with the correct transmission level for ordinary telephone sets). Further, the *FW-1* system permits the use of identical circuits for 2-wire and 4-wire offices, simply omitting in the former case the additional equipment needed for 4-wire operation.

Toll Offices With 2-Wire Switching

Before the development of toll offices employing 4-wire switching, toll calls were extended at intermediate as well as terminating points by using the two wires of the talking leads only. At intermediate points the line terminals of the two hybrids were interconnected over two wires, and the network terminals were closed by compromise networks. At end points the line output of the terminating hybrid was extended over two wires to the local connecting trunk. Due to the relatively low terminal-return loss of the hybrid circuit, the inherent gain of modern 4-wire circuits could not be fully used, and an artificial loss had to be inserted between the hybrid output and the connecting trunk. Since in a through-switch state somewhat lower losses are permissible than in the terminating state, the artificial loss was removed for through connections in 2-wire offices which were arranged for pad switching, and multiswitch connections were established at VNL.

Nevertheless, each intermediate 2-wire office in a multilink connection represented a potential source of echo and instability, due to the low return loss at each hybrid circuit, whereas, in a 4-wire intermediate office, echo paths are automatically eliminated. Therefore, 2-wire offices could not be substituted for 4-wire offices, without an increase of the VNL, unless the return loss at each hybrid point could substantially be increased.

Low return loss is due to unbalance between the office equipment and wiring and the balancing network in the hybrid circuit. The average return loss of a typical 2-wire office, free of equipment and wiring errors, is about 14 db, and unless the average return loss can be raised to at least 27 db, trunks terminating in a 2-wire switching office must be operated at a higher VNL than when there are 4-wire offices at both ends. To achieve such high value of return loss in 2-wire offices, the length of inside cable must be limited, large-gauge cable conductors used, the capacitances of different cable conductors adjusted to a uniform value, and the compromise networks compensated for the impedance of the office cabling and equipment.

Four-wire switching equipment using network switching arrangement automatically eliminates return loss at a relatively small increase in equipment over 2-wire switching. Even for smaller offices it is advisable to investigate whether it is not preferable to provide 4-wire switching rather than expend considerable amount of labor and material needed to improve the return loss in 2-wire offices to a point where their stability may approach that of a 4-wire office. Further, it should also be considered that at each expansion of the office, the new equipment must similarly be balanced.

It has been shown that in a 4-wire office, part of the gain available in the 4-wire circuit may be transferred to passive 2-wire lines or high-loss connecting trunks, thereby improving the over-all transmission of many toll connections. A transmission improvement on high-loss trunks may also be obtained in a 2-wire office, if the switching pad, ordinarily only used for local terminating calls, is eliminated entirely. The removal of the switching pad and its control equipment may also reduce the cost of such toll switching offices. However, assuming that TNL = VNL + 4 as before, some substitute must be found to replace the 2-db losses at each end, which were previously supplied by the switching pads.

Two mutually exclusive conditions may now be considered. The first case involves connecting trunks to local offices in the same building as the toll offices (or in the immediate vicinity) so that the trunk loss is less than 2 db. The second case involves connecting trunks having more than 2-db loss. Now, if a 2-db fixed pad is added to every trunk which belongs to the first group, the transmission status of this group remains exactly the same as it was with the previous switching pad arrangement. Since the transmission loss of a trunk in the second group is more than 2 db, no substitute for the switching pad is needed. Transmission over these trunks is improved by 2 db at each end; unfortunately, by the omission of the 2-db pad, the echo and singing margins are also lowered by 4 db at each end. If these conditions are to remain substantially unimpaired with the omission of the switching pad, the return loss must be improved by at least 4 db. Impedance compensation of the connecting trunks and capacitance adjustment of the office cabling is used to bring this about.

It has been shown previously that means are incorporated in a 4-wire system to transfer gain to high-loss connecting trunks; and since pads are included between the hybrid and the connecting trunk, as well as between the hybrid and the compromise network, adequate terminal return loss can be obtained in the 4-wire system without any further artifice of impedance compensation and capacitance adjustment.

Summary

A new 4-wire switching telephone system has been developed, incorporating the tail-eating, or network switching, principle of operation in its transmission paths. Four-wire switching of telephone calls has the advantage over 2-wire operation that a very high degree of stability and freedom from echo are automatically obtainable on built-up connections, whereas with 2-wire switching a considerable amount of balancing and impedance compensation within the central office must be carried out to approach the inherent balance of a 4-wire office. The excess gain available in a 4-wire through connection permits a flexible system of pad switching, by means of which some of the gain may also be transferred to high-loss local connecting trunks. In 2-wire offices this result is obtainable only by additional balancing, by providing built-in pads in connecting trunks, and by impedance compensation of connecting trunks.

The tail-eating method of 4-wire switching also has important advantages over methods which separate directions of transmission. It eliminates the need for hybrid circuits in local connecting trunks; it permits simplification of the switching circuits; it permits the use of the same circuits for 4-wire as well as 2-wire switching systems; it automatically provides 2-wire operation in every case where the 4-wire control conditions are not established; it permits the use of compromise balances and provides for an efficient system of pad switching while still maintaining equal transmission-power levels in both paths through the switching equipment; and it permits operators or test personnel to intervene, using conventional telephone circuits.

Discussion

J. W. Emling (Bell Telephone Laboratories, Inc., New York, N. Y.): Dr. Molnar presents an interesting discussion of the transmission problems of toll switching but the paper would have been enhanced by references to prior publications on the subject, particularly since much of the material has been so adequately covered by elementary texts on communications and in papers.[1–6]

It would be unfortunate if the impression were gained from the title that the network switching scheme is new. While the application is undoubtedly new in some of its details, the basic scheme is not particularly novel. For example, a patent granted in 1923 describes the "line-to-line and net-to-net," or tail-eating, method of interconnection in some detail.[7] The same general scheme has also been used in some European countries for a number of years.

The advantages of 4-wire switching as compared to 2-wire are well known. The principal item of interest raised by the paper concerns the merits of different types of 4-wire switching. As indicated by the author, the directional switching and network switching arrangements are in some ways alike but also each has its own peculiar advantages. In Bell System applications of 4-wire switching, the directional switching method has been found preferable because of the complete absence of reflection on through switches, the smaller number of hybrid coils required, and the facility with which precision networks can be associated with 2-wire connecting trunks.

Under other circumstances, it is possible that the network switching method would be advantageous although the bases for some of the claimed advantages are not altogether obvious. In this connection it is to be regretted that the author did not explain further the reason why a 30-db return loss is obtained with one of the wires completely open.

REFERENCES

1. CROSSBAR TOLL SWITCHING SYSTEM, **L. G. Abraham, A. J. Busch, F. F. Shipley.** *AIEE Transactions*, vol. 63, June 1944, pp. 302–09.

2. AUTOMATIC SWITCHING FOR NATION-WIDE TELEPHONE SERVICE, **A. B. Clark, H. S. Osborne.** *Ibid.*, vol. 71, pt. I, Sept. 1952, pp. 245–48.

3. FUNDAMENTAL PLANS FOR TOLL TELEPHONE PLANT, **J. H. Pilliod.** *Ibid.*, pp. 248–56.

4. NATION-WIDE NUMBERING PLAN, **W. H. Nunn.** *Ibid.*, pp. 257–60.

5. AUTOMATIC TOLL SWITCHING SYSTEMS, **F. F. Shipley.** *Ibid.*, pp. 261–69.

6. TRANSMISSION DESIGN OF INTERTOLL TELEPHONE TRUNKS, **H. R. Huntley,** *Ibid.*, vol. 72, Nov. 1953, pp. 670–76.

7. **George Crisson, A. F. Rose.** *U.S. Patent no. 1,465,298*, Aug. 21, 1923.

Imre Molnar: The first paper quoted by Mr. Emling (reference 1) describes, in 1944, a crossbar toll switching system which is now considered as obsolete. There is about one page devoted to the transmission features of this system, which has no similarity and is irrelevant to the subject matter of my paper.

The second set of papers quoted (references 2–5 of the discussion) describes the nation-wide toll switching plan and a more contemporary crossbar toll switching system. The subject of transmission is likewise contained within about one page and is restricted to generalities.

The Huntley paper (see reference 6 of the discussion) gives an excellent outline of the transmission problems of the toll trunk plans, while the transmission problems treated in my paper concern mainly central-office switching equipment designed around the tail-eating principle. Mr. Huntley was not concerned with 4-wire switching at all, and only in one-half column he summarizes the problems encountered with 2-wire switching. These problems, as a side issue, were also further elaborated and brought up to date in my paper.

Personally, I could not find any elementary or advanced text covering the principle of 4-wire switching in general or the tail-eating principle in particular, other than a German text describing the directional 4-

wire switching systems used by German manufacturers, and in a house publication of limited circulation by Siemens Brothers of England.

I would regret it if anyone would infer from the title of my paper that the principle of network switching is new, which it is certainly not. However, my application of this principle as a substitute for directional 4-wire switching within the framework of the transmission plan adopted for North America, as well as the complete switching system in all of its details, which accomplishes this, is new. I believe that this is sufficient to qualify the subject matter of my paper as a recent development, just as the perennial papers on new developments in crossbar switching systems are definitely of general interest even though the principle of crossbar switching considerably predates the principle of network switching. Incidently, to the best of my knowledge tail-eating systems have not been employed in Europe. There are installations in operation in Australia, South Africa, and Canada which employ tail-eating 4-wire switching, but the transmission objective and the switching principles are very different from those described in my paper.

While there may be some specific applications where directional switching would be preferable, for general applications the advantages claimed for the latter do not seem to be valid. In a properly designed network switching system the reflections on through switching are absent for all practical purposes. The number of hybrid coils is very considerably less with the tail-eating scheme. In an actual application of a toll switching office with about 3,000 trunk terminations the number of hybrids was reduced by over 50 per cent as compared with the directional switching scheme. As to the facility by which precision networks can be associated with 2-wire connecting trunks, Fig. 10 and its description in the text demonstrates that it can with equal facility be done by this new switching system.

Mr. Emling concurs with me that 4-wire switching is superior to 2-wire switching, but unfortunately there is also a widespread impression in the telephone field that 2-wire switching can be made equivalent, with reasonable effort and at a reasonable cost to 4-wire switching in almost any application. I hope that my paper will help to clarify this, and at the same time point out the method by which a system superior to 2-wire switching can be obtained without the cost penalty of more conventional 4-wire switching methods.

The conditions with one of the wires open can be visualized by adding the symmetrical half of the windings (as in Fig. 1) to Fig. 5, and opening up, say, the upper line lead between the no. *1* windings of the east and west hybrids. Assuming that the west terminal is the sending end, practically all energy will now flow to A_E through no. *2* west winding, the continuing net lead and the intact line lead. A small amount of energy will also leak through to A_W because there will be a difference in the respective potentials across the no. *1* and *2* coils of the west hybrid, due to the difference in the nonreactive voltage drops through these windings. The power ratio between A_W and A_E is in the order of 30 db, but this value will vary in practice, dependent on the design of the hybrid and the degree of symmetry of the elements used in the connection.

The Fused Silicon Rectifier

HERBERT W. HENKELS
ASSOCIATE MEMBER AIEE

IN THE past year or two, the culmination of intensive efforts in the development of silicon materials and devices has been the introduction of practical silicon rectifiers to the electrical industry. The widespread use of such rectifiers may be expected in the coming years. At the moment, there appears to be no competitor in the range of operating temperatures below 200 C (degrees centigrade).

Silicon rectifiers can be made by techniques characterized by: (1) a diffusion operation; (2) an evaporation operation; (3) a junction-pulling operation; and (4) a fusion operation. The fusion technique of making rectifiers can be further divided into that employing a wire electrode in the operation (fused-wire diodes) or that using flat solder preforms. It is possible that silicon rectifiers can be produced by a few additional methods, but these are in the research stage of development. All of the four methods of producing rectifiers have been tried successfully; all have certain advantages and disadvantages. Thus, method 1 makes the volume production of *p-n* junctions easy, but subsequent sectioning, etching, and lead attachment somewhat troublesome. Method 2 again permits the large-scale production of *p-n* junctions, but etching and lead attachment are somewhat difficult. The third method is capable of producing the most perfect *p-n* junctions with resulting high-voltage characteristics, but slicing and sectioning problems can cause rather poor forward-drop characteristics. The ultimate limit on the degree of perfection of *p-n* junctions made in large-scale production by the fusion method is lower than that of pulled *p-n* junctions and probably lower than that which may be attained at some future date in diffused junctions. However, very good fused junctions can be prepared and other processing can be made quite simple so that from an over-all economic standpoint the fused junction rectifier is very attractive. Both the fusion techniques employing wires or solder preforms are discussed in this paper.

There are two principal theories of rectifier operation; one based upon a model in which the asymmetry in resistance characteristic arises from the differences in magnitudes of quasi-equilibrium field and diffusion currents, in two directions of conduction, and the other based on a model in which the asymmetry arises from differences in nonequilibrium diffusion currents in both forward and reverse directions.

The first theory, a modification of that enunciated in somewhat different forms by Mott,[1] Schottky and Spenke,[2] and Davydov,[3] describes a field current which is limited by the electron depletion of a dipole region between a semiconductor and counterelectrode contact. The original theory was designed for metal semiconductor contacts and has not been employed in the silicon and germanium *p-n* junction cases of recent interest. However, if the field-limited case is considered for a junction of equal carrier density *p*- and *n*-type materials, Appendix I, the reverse current depends as given in equation 1 upon the semiconductor properties.

$$i=6.19\times10^{-28}n_i(\mu_p+\mu_n)(N_A)^{1/2}(V_{DT}+V)^{1/2} \quad (1)$$

where

i = current density, amperes per square centimeter (cm²)
n_i = intrinsic carrier number density, 1 per cubic centimeter (cm³)
μ_p, μ_n = mobilities of holes and electrons, respectively, cm² per volt-second
N_A = acceptor (or donor) number density, 1 per cm³
V_{DT} = inner contact potential, volts
V = applied voltage, volts

The second theory, that of Shockley,[4] describes a reverse current limited by a

---

Paper 56-975, recommended by the AIEE Metallic Rectifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted June 14, 1956; made available for printing August 21, 1956.

HERBERT W. HENKELS is with Westinghouse Electric Corporation, Youngwood, Pa.

The author acknowledges the contributions of J. Lasco, H. Camp, T. Lawson, J. Pappas, and J. Fortier, and expresses thanks to Dr. S. J. Angello, under whose guidance the program was carried out.

diffusion current in the bulk of the semiconductor near the *p-n* junction and depends on the semiconductor properties and constants as given in equation 2.

$$i=2.37\times10^{-27}n_i^2T^{1/2}\left[\frac{\mu_p^{1/2}\mu_n\rho_n}{\tau_p^{1/2}}+\frac{\mu_n^{1/2}\mu_p\rho_p}{\tau_n^{1/2}}\right] \quad (2)$$

with additional symbols defined:

T = absolute temperature, degrees Kelvin (K)

τ_n, τ_p = lifetimes of electrons and holes, respectively, microseconds (μsec)

ρ_n, ρ_p = resistivity of *n*- and *p*-type materials, respectively, ohm-centimeters (ohm-cm)

The quantity n_i is obtained from equation 3

$$n_i^2=AT^3\exp\left(\frac{-E_G}{kT}\right) \quad (3)$$

where previously undefined symbols are

E_G = intrinsic energy gap in the semiconductor, electron-volts

k = Boltzmann's constant, electron-volts per K

A = factor describing density of states in conduction and valence bands (see Table XI)

The current predicted by equation 1 is much larger than that by equation 2. The former equation should hold only when lifetimes are very small, the latter when lifetimes are such that the diffusion lengths defined as $(D\tau)^{1/2}$ or $(kT\mu\tau)^{1/2}$, where D is the diffusion constant of the carrier under consideration, are considerably larger than the extension of the dipole region at the *p-n* junction. The general expression for current through a *p-n* junction with arbitrary values of τ in the semiconductors has not been worked out.

Both theories are concerned only with rectification across a contact between infinite planes of semiconductors; consequently, in the theoretical expressions the potential within the semiconductors in the presence of the dipole at the *p-n* junction is simplified and edge-surface effects are neglected. Since the extension of the counterelectrode junction is much larger than the thickness of the dipole region, there is no appreciable error in the method of calculating potential. However, the parameters in the junctions may have values at the surfaces different from those in the bulk of the semiconductor. Variations in mobilities and resistivities may be neglected in comparison with variations of E_G and τ. Important effects on resistivity of surface changes in E_G have not been noted in bulk silicon semiconductors, but the value of τ at silicon surfaces may vary radically from that in the bulk materials. The principal reason for including the quasi-equilibrium junction theory, equation 1, is to show the limiting case of very small lifetimes.

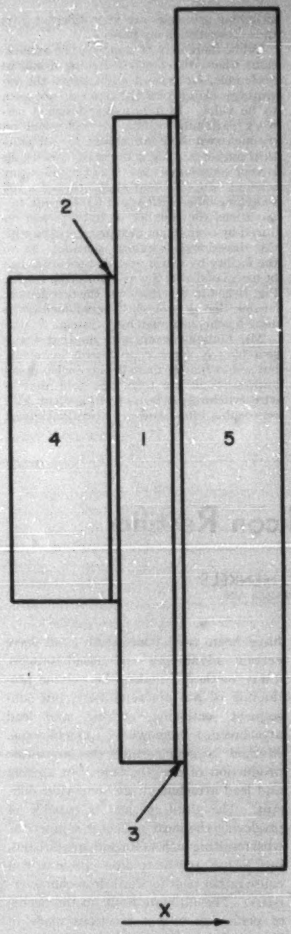


Fig. 1. Rectifier cell components

The quasi-equilibrium theory predicts a forward current which varies as given in equation 4 at low voltage and approaches asymptotically a linear dependence on voltage at high voltage, the slope depending on the bulk conductance of the semiconductor.

$$i=\frac{6.19\times10^{-23}n_i(\mu_p+\mu_n)(N_A)^{1/2}(V_{DT}-V)^{1/2}\left(\exp^{\frac{V}{2T}(1.16\times10^4)}-1\right)}{1-\exp\left(\frac{-(V_{DT}-V)}{2T}(1.16\times10^4)\right)} \quad (4)$$

The nonequilibrium theory for the forward characteristics predicts a current density given by equation 5 at low voltages (to ~0.1 volt), one which depends on the same factors at somewhat higher voltages with the value of the exponent changing to $(1/2)(V/kT)$.

$$i=2.37\times10^{-27}n_i^2T^{1/2}\left[\frac{\mu_p^{1/2}\mu_n}{\tau_p^{1/2}}\rho_n+\frac{\mu_n^{1/2}\mu_p}{\tau_n^{1/2}}\rho_p\right]\left(\exp^{\frac{V}{T}(1.16\times10^4)}-1\right) \quad (5)$$

At high forward voltages, the current increases directly with the voltage with the slope depending upon the conductance of the bulk material as modulated by carrier injection. The general forward characteristic depends upon consideration of both field and diffusion currents and has not been worked out.

The question of the mechanism of the breakdown of silicon *p-n* junctions has not been completely resolved although the avalanche mechanism seems probable. Recent empirical results have related breakdown voltages (V_b) to resistivities as given in equations 6.

$$V_{b_n}=86\rho_p^{0.64}$$
$$V_{b_p}=23\rho_p^{0.75} \quad (6)$$

In the process of fusing the counterelectrode, a recrystallized, usually highly doped, region of material of opposite conductivity type to that of the starting blank is produced. Thus, the properties of the recrystallized layers usually drop from the equations.

With values of parameters (which may be assumed uniform in the material used for making silicon rectifiers) given in the literature, $E_G=1.21eV$, $\mu_p=490$ cm² per volt-second, $\mu_n=1{,}100$ cm² per volt-second, the theories predict rectification with any values of lifetime or resistivities. Dice of silicon with fairly uniform resistivities can be obtained. Furthermore, in the range of 10 ohm-cm the resistivities do not change appreciably in the processing to be described so that measurements of resistivities before fusion are meaningful. Very-high-resistivity material may change its resistivity, or even type, on heating. Prediction of its behavior depends on a detailed knowledge of the crystal impurities and defects.

Unfortunately, the lifetime across the surface (y z plane in Fig. 1) of a *p-n* junction cannot be measured. This lifetime may be expected to vary in fused

Table I. Properties of Pure Metals as Counterelectrode Solders

Metal	Liquidus of Si Eutectic, C	Wet Si and Counter-electrode	Solid During Processing	Solid, Mechanically, and Chemically Stable In Use	Vapor Pressure	Diffusible Traps*	Thermal Conductivity,† Centimeter-Gram-Second Units	Modulus of Elasticity, $\times 10^6$ Psi	Coefficient of Thermal Expansion, $\times 10^{-6}$ Inch Per Inch C	Product Modulus of Elasticity and Difference of Thermal Expansion, Alloy-Silicon, (Si $\sim 4\times 10^{-6}$)	Solidus Temperature, C	Doping Effect
Al	577	~800 C yes	yes	quite brittle	low	no	0.37	10.3	21	175	577	p^+
Cu	805	yes	yes	quite brittle	low	no	0.94	16	16.5	200	805	
Ag	830	~900 C yes	yes	quite brittle	low	no	1.0	11	19.7	173	830	n^-
Au	370	~700 C yes	borderline	quite brittle	low	no	0.71	12	14.2	122	370	n^-
Mg	645	yes	yes	very brittle	low	no						
Sb	630.5	low solubility	yes	quite brittle	high	no	0.045					

*Those having a noticeable influence on silicon rectifier properties.

†Thermal conductivities, moduli of elasticity, and coefficients of thermal expansion are those of the elements, except for the case of the Al-Si eutectic.

junctions for two principal reasons: (1) defects in the original crystal blank; (2) internal unwet areas. The surfaces of internal unwet areas should be coated with thin films of the electrode solders so that the lifetimes at the peripheries of these regions may be expected to be very low. If the recrystallized region is sufficiently imperfect or of very small extention, as occurs, for example, when recrystallization is not carried out correctly, the properties of the recrystallized layer may no longer be neglected. In such cases, the complete equations must be used. It is seen from equation 2 that the total current will be increased as the lifetime in the recrystallized layer is decreased. If the recrystallized layer thickness (x) is less than the diffusion length (L), the diffusion current in the recrystallized region is increased by the factor L/x.

An equally serious problem in measuring lifetimes in fused junctions in that associated with determining the lifetime in the regions of the exposed edges of the p-n junctions. An additional problem arises in relating surface and bulk lifetimes after fusion to the bulk lifetime measured before fusion and processing parameters. In the usual process, the bulk lifetime of silicon decreases considerably.

Thus, some success has been achieved in relating breakdowns to resistivities since reliable resistivities can be measured and sufficiently perfect junctions made to maintain surface leakage currents small enough as to not interfere with estimation of the breakdown point. On the other hand, it has proved to be more difficult to relate theoretical and experimental currents since exact knowledge of bulk and surface lifetimes in the finished junction is required. The characteristics of finished junctions and rectifiers are discussed in the section on cell and rectifier properties. For the moment, the problem is the intelligent selection of silicon cell components to give useful rectifiers.

Basic Cell Fabrication

Choice of Cell Components

Any rectifier cell consists of the parts shown in Fig. 1. The parts in the figure are discussed in order of their importance to the finished device; i.e., (1) silicon, (2) counterelectrode solder, (3) base solder, (4) counterelectrode, and (5) base electrode.

1. *Silicon*

The resistivity of silicon for very-small-area junctions may be selected to give breakdowns distributing near the predicted values, i.e., for 100-volt breakdown about 2- or 3-ohm-cm material is required. In larger areas cells, breakdowns are not as sharp and the resistivity must be selected on the basis of empirical results to give the average value of the distribution of "breakdowns" desired. With n-type silicon about 10-ohm-cm material and with p-type silicon 30 ohm-cm is suitable for the manufacture of cells with breakdowns distributing over 50 volts to 300 or 400 volts.

Silicon crystals used were grown from du Pont silicon in both needle and compact forms. Crystals were grown in vertical pulling furnaces with quartz crucibles in graphite holders. Slow rotation was employed in pulling. A carefully cleaned inert gas of helium was employed when doping silicon n-type with arsenic. Crystals doped n-type with gold and silver have been grown in vacuum. P-type crystals have been grown in vacuum without any added doping agent, residual boron contributing the acceptor levels.

Average lifetime in ingots of silicon of about 20 μsec is suitable for the production of silicon rectifiers under discussion. Average lifetime in the finished junction is in the range 1 to 5 μsec.

Crystals with grain boundaries or excessive lineage do not give results as good as those from perfect single crystals. In production, it is difficult to detect any bad effect of twinning, however.

No effects of electrical anisotropy have been considered in the theoretical results and none have been observed experimentally. Crystals grown in the (111) and (100) directions have been used with equal success. There are probably small differences in properties which were not detected with the random error residual present.

The listing of some of the mechanical properties of silicon is essential to the later discussion of counterelectrodes, counterelectrode solders, base solders, and bases. The silicon must be sufficiently strong to withstand mechanical stresses occurring in manufacture and application. The applications will include military ones so that, for general use, the silicon must be capable of withstanding several conditions of shock, mechanical vibration, temperature variation, etc., imposed by military specifications, for example, *MIL-E-5272-A*.

The thermal expansion coefficient was measured in the (111) direction and perpendicular thereto in the range of temperatures from −60 C to 800 C. Samples were of sizes 1/4 by 1/4 by 2 inches and 1/4 by 1/4 by 1 inch. The thermal coefficients of expansion in the (111) direc-

Table II. Binary Counterelectrode Alloys

Alloy	Solidus of Binary System	Notes
Al-Ag	~558	Addition of Ag to Al may lower solidus, but the effect is small
Al-Cu	~510 (of ternary)	Addition of 3–4% Cu may lower solidus to range approximated, but result is little improvement beyond pure Al
Al-Mg	~451	Addition of Mg to Al lowers the solidus. The modulus will change little
Al-Ga	~425	Addition of Ga lowers solidus to the indicated range. The modulus will decrease. The resulting alloy is very ductile
Ag-Al	~780	Addition of Al to about 5% will lower solidus considerably and may provide *p*-type conductivity
Ag-In	~693	Additions of In to about 20% will lower the solidus considerably. The modulus may be expected to decrease
Ag-Ga	~724	Addition of Ga to 12% will lower the solidus and the modulus
Cu-Al	~1020	Al bronzes will not appreciably reduce solidus temperatures
Cu-Ga	~909	Solidus should be reduced as well as the bulk modulus of the ternary
Cu-In	~715	Solidus will be reduced considerably. The effect on the modulus is unknown
Au-Al	~417	Addition of Al can lower the solidus to the range indicated. The Au-Al alloys are extremely brittle and cannot be rolled
Au-In Au-Ga	?	Small amounts of In and Ga dissolve in Au, but phase diagrams are not available. The resulting alloys with Ga dope silicon *p*-type as desired. The low modulus of Au and its ductility are not adversely affected

tion were 4.23, 4.16, and 4.26, by 10^{-6} inch per inch C in three trials above room temperature; in the perpendicular direction 3.66, 3.58, 3.85, by 10^{-6} inch per inch C in three trials. Averaged below room temperature to −60 C, the coefficients were approximately 2.0 by 10^{-6} inch per inch C in both directions.

The modulus of elasticity and ultimate stress in silicon were measured. With the small specimens used, a variety of results for ultimate stress were obtained from 12,650 to 21,550 psi (pounds per square inch). The variation resulted from the irregularities in shape of specimens that occurred during cutting, lapping, and etching. Values of Young modulus were more consistent and varied from 22.6 to 32.5 × 10^6 psi. The ultimate elongation is seen to be very small, approximating 0.7 mil per inch. Consequently, the mechanical stresses must be small to prevent cracking of the silicon during preparation or operation. Moreover, the electrical properties and dependence of these on temperature cycling are related to the strain in the silicon lattice.

Besides having appropriate electrical and mechanical properties, the silicon semiconductor must be stable under conditions of long operation of the rectifier. Diffusion coefficients of impurities in silicon are small so that at temperatures of 200 C the rates of diffusion of donor and acceptor impurities are low. Thus, bulk diffusion does not seem to be important in the alteration of the silicon properties with period of operation of rectifiers, as is true in the selenium rectifier.

The thickness of the wafer of silicon used depends upon the penetrations of counterelectrode and base solders in the fusion operation. In well-prepared rectifiers, the thickness of these solders is dictated by mechanical factors discussed later so that the desired thickness of silicon can be calculated from consideration of the diffusion length of injected carriers in the finished rectifier cells. However, practical factors of control of the fusion operation place a lower limit on thickness for optimum yields of good cells. This thickness is in the range of 8–10 mils for the silicon in the process under discussion.

The silicon surface must be carefully lapped and etched after slicing in order to present a clean, uniform surface to the counterelectrode alloy. Successful rectifier cells can be made with blanks etched to a smooth specular surface or to a satin finish. The latter surface results when very little material is removed after lapping.

2. *Counterelectrode Alloy*

In the fused silicon rectifier, the *p-n* junction is produced by melting an appropriate alloy in contact with the prepared silicon blank. On cooling, silicon, dissolved in the alloy, redeposits on the base silicon forming a layer of opposite conductivity type to that of the bulk and producing the *p-n* junction. The exact choice of alloy dictates the fusion temperature. However, in all cases, the recrystallization must be carried out in such a manner that the temperature decreases from the alloy to the silicon in order to obtain the necessary thickness of redeposited silicon, free of metal inclusions and defects, of reasonable lifetimes. It has been found possible to use rapid cooling rates of approximately 50–100 C per minute in the range of redeposition, provided the correct temperature gradient is employed.

The counterelectrode alloy, besides producing a silicon layer of the desired conductivity type, must fulfill a number of other conditions:

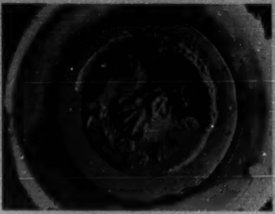


Fig. 2. Cracked silicon cell

(*a*). It must be liquid at a temperature as far as possible below the melting temperature of silicon or the other components of the cell.

(*b*). It must wet the silicon and counterelectrode.

(*c*). It should be solid at all temperatures encountered during processing after the fusion cycle and solid, mechanically, and chemically stable during storage or operation (for safety 400 to 500 C).

(*d*). It should have a low vapor pressure at all temperatures encountered in the fusion cycle.

(*e*). It should not contain diffusible traps.

(*f*). It should have high thermal conductivity.

(*g*). It should not produce excessive strain in the silicon.

Table III. Binary Alloys for Counterelectrode Solders on P-Type Material or Base Solders on N-Type Material

Alloy	Remarks
Au-Sb	Small amounts, 0.01 to 0.1% (per cent), of Sb in Au produce an alloy which retains the properties of pure Au and which dopes silicon n^+. Larger amounts of Sb produce more brittle alloys which are difficult to roll and which crack upon temperature cycling
Au-As	The alloy is similar to the Sb alloy. As is somewhat less convenient to work with
Cu-Sb	Additions of Sb up to about 5 wt (weight) % will dissolve in Cu and lower the solidus to about 900 C. The solidus of the ternary with silicon may be expected to be lower, but not sufficiently so as to make the alloy interesting
Cu-As	Additions of As to about 7 or 8 wt % will dissolve in Cu. The solidus is lowered to 689 degrees but the liquidus remains quite high (980 to 990 C). The silicon ternary will also have a high liquidus
Ag-Sb	Additions of Sb to about 7 wt % will dissolve in Ag and lower the solidus to about 780 C, the liquidus to about 900 C. The Ag-Sb alloys are less brittle than the corresponding Au-Sb alloys
Ag-As	Additions of $4\frac{1}{2}$ wt % As will lower the solidus to 595 C, the liquidus to about 880 C. The ternary with silicon should have a still lower solidus and liquidus temperature

Table IV. Alloys of Ag with Lower Solidus Temperatures than Pure Ag

Alloy	Solidus Temperature, C	Element Added
Ag-Cd	730	36.5 wt % Cd
Ag-Cu	779	8.8 wt % Cu
Ag-Ge	640	4.5 wt % Ge
Ag-Pb	304	about 4% Pb
Ag-Sn	724	13.3 wt % Sn
Ag-Zn	710	22 wt % Zn

The first condition arises from the desire to retain the original lifetime characteristics as much as possible. It is not absolutely necessary that the counterelectrode solder remain solid during processing and good cells have been made by a process in which the condition was violated, but a solid electrode simplifies process controls. The fourth condition is desirable from the standpoint of the ease of post-etching and again is not essential. Condition *e* expresses the desire to retain lifetimes during processing steps which take place at high temperatures where diffusion is rapid. The temperature drop in the solders should be low to permit the greatest possible heat dissipation within the cell, as discussed in the section on encapsulation and heat dissipation. The electrical properties depend upon strain to some extent, but more important, the silicon must be safely below its ultimate elongation in the range of temperatures found in the processing of the basic diode and application of finished rectifiers.

Condition *g* can be re-expressed in terms of the properties of the counterelectrode solder and silicon by observing that the strain in the silicon (part *1* in Fig. 1) is given by the following, see Appendix II:

$$\frac{S_3}{E_3}=\left\{\frac{\sum_{n=3\text{-}5}(\alpha_n-\alpha_3)t_nE_n(T-T_4)+\sum_{n=1,2}(\alpha_n-\alpha_3)t_nE_n(T-T_2)}{\sum_{n=1\text{-}5}E_nt_n}+\frac{(T_2-T_4)\sum_{n=1,2}E_nt_n\sum_{n=3\text{-}5}E_nt_n(\alpha_n-\alpha_3)}{\sum_{n=1\text{-}5}E_nt_n\sum_{n=1\text{-}5}E_nt_n}\right\} \tag{7}$$

where

S = stress in the silicon
E_n = modulus of elasticity of element n
α_n = coefficient of thermal expansion of element n
T_4, T_2 = solidus temperatures of base and counterelectrode solders
t_n = thickness of element n

It is noted that the minimum stress in silicon is obtained with a low solidus temperature for the counterelectrode alloy, a low-bulk modulus, a small thickness of the alloy, and a coefficient of thermal expansion matching that of silicon. Thus, condition *f* is restated

(*h*). The product of thickness, modulus of elasticity, and difference of thermal coefficients of silicon and the alloy must be as small as possible.

(*i*). The solidus temperature of the alloy should be as low as possible consistent with condition *c*.

The eight conditions should be considered before noting the doping effect of the alloy since the addition of trace elements may adequately dope the redeposited silicon layer without affecting the listed properties of the alloy. The first four conditions limit the useable pure elements to silver (Ag), gold (Au), aluminum (Al), and copper (Cu). Antimony (Sb) and magnesium (Mg) have the correct melting ranges and are expected to wet the silicon and counterelectrode, but both should form very brittle alloys with silicon. In addition, the vapor pressure of Sb is too high for use in vacuum fusion. The discussed properties of the useful pure metals are listed in Table I. The metals noted all have good thermal conductivities, Al having the poorest. Of the four alloys, Cu seems least attractive, having the highest product of bulk modulus and difference of expansion coefficients and a high solidus. The solidus temperatures of Ag and Cu are too high, that of Al somewhat higher than the set range to 500 C. The solidus of the Au-silicon alloy is somewhat lower than desired, but the other properties are so attractive that the discussion will retain Au and its alloys.

To produce the correct *p*-type of redeposited silicon (against *n*-type bulk silicon), acceptors must be present in alloys of Cu, Ag, and Au. Thus, the binaries of these elements with gallium (Ga), indium (In), or Al should be considered. To produce the desirable heavily doped *n*-region in contact with *p*-type bulk silicon, donor elements must be added: phosphorus (P), arsenic (As), Sb. Of these, Sb is the least difficult to handle and is less volatile in the fusion cycle. Thus, binaries of Sb and Ag, Cu or Au are interesting.

To summarize, it is desirable to reduce the solidus of Ag-silicon, Cu-silicon, and Al-silicon alloys and the modulus of elasticity by addition of another element or elements. It is also necessary to add an acceptor element to Ag and Cu alloys for use as counterelectrode solders against *n*-type silicon, and to add a donor element Sb to the same alloys for use as counterelectrode solders aganst *p*-type silicon.

Table V. Ternaries of Ag, Pb, Sb

Ternary	Liquidus, C	Solidus, C	Ductility
Ag—1% Sb—1.5% Pb	910	555	soft
Ag—1% Sb—2 % Pb	~907	525	soft
Ag—1% Sb—3 % Pb	904	490	soft
Ag—1% Sb—4 % Pb	~890		soft
Ag—1% Sb—5 % Pb		415	soft

Table II lists possible binaries for use as counterelectrode solders against *n*-type silicon. Reviewing the tabulated items, it appears that additions of Ga and Mg may noticeably reduce the solidus temperature of an Al counterelectrode solder. The modulus of a Mg-Al system will change little whereas that of the Ga-Al system will be reduced. The Ga-Al alloys are very ductile.

It appears that the addition of Al to Cu and Ag will not produce electrodes superior to those of pure Al. Although the solidus temperature of Au is reduced markedly with small amounts of Al, the resulting intermetallics formed produce a very brittle alloy which cannot be rolled. Alloys of Ga with Ag and Cu have not been investigated by the author. These may be expected to have lower solidus temperature and bulk moduli than those of the pure metal-silicon alloy, but it is doubtful that the properties will be superior to those of Al-silicon or Al-Ga-silicon alloys. The Au-Ga system dopes silicon *p*-type and has all the desirable properties of the Au counterelectrode alloy.

In order to control electrode penetration, Al-silicon eutectic alloy has been employed instead of the Al element.

Al or Al-silicon eutectic alloy has been used as counterelectrode material in almost all of the experiments with *n*-type bulk silicon. Al-Ga alloys have been studied as a possible improvement over these. However, the Al and Al-silicon counterelectrode materials have proved to be satisfactory and capable of producing rectifiers in which excessive strains do not occur during temperature cycling from −55 to 200 C. In the development of small communication diodes (by the fused-wire technique), Au-Ga alloys have been employed. However, a new type of small-area diode has been developed which is more promising than the fused-wire type. The new diode employs the same counterelectrode solders as the large power types.

The counterelectrode solder thickness is approximately 2 mils, this being sufficiently small to prevent excessive strain

Table VI. Properties of Counterelectrode Materials

Material	Elastic Modulus, Psi $\times 10^6$	Coefficient of Thermal Expansion, Inch Per Inch C $\times 10^{-6}$	Thermal Conductivity, Calories Per Second (Cm^2) (C Per Cm)
Mo	50	5–6*	0.35
Mo, Kovar-alloy-clad	50	5–6*	0.35
Mo, Ni-clad	50	5–6*	0.35
Kovar alloy	19.5	4 to 445 C >5 at high temperature	0.03
Zi	11	5	0.04
W	50	4.3	0.48
Ta	27	6.5	0.13
Invar alloy (36% Ni, Fe)	21	2.7 to 200 C 11.5 to 200–315 C 15.5 to 315–425 C	0.025
Hypernik alloy (47–50% Ni, Fe)	24	5.5 in small range	0.029
Cr	36	6.2	0.16

*See Table VII.

Table VII. Average Coefficient of Thermal Expansion of Mo Sheets (Inch Per Inch C $\times 10^{-6}$)

Temperature Range, C	Plain Mo Sheet	Mo Sheet, Kovar-Clad One Side	Mo Sheet, Kovar-Clad Two Sides	Mo Sheet, Ni-Clad One Side and Stainless-Steel-Clad One Side
100–200	5.18	5.60	5.00	5.58
100–300	5.27	5.63	5.28	5.67
100–400	5.43	5.67	5.37	5.73
100–500	5.55	5.74	5.75	5.75
100–600	5.63	5.83	5.87	5.77
100–700	5.69	5.93	5.95	5.77
100–800	5.73	5.99	6.01	5.78
100–900	5.75	6.04	6.06	5.78

Table VIII. Parts Employed in Communication Diode and Larger Rectifier Sides

	N-Type Communication Diode	N-Type Silicon Power Rectifier	P-Type Silicon Power Rectifier
Counterelectrode	Ta	Mo, Ni-plated one side	Mo, Ni-plated one side
Counterelectrode solder	Al-silicon eutectic	Al-silicon eutectic	Au-1% Sb
Silicon	Chosen on basis of breakdown desired by formula	~10 ohm-cm	±30 ohm-cm
Base solder	Ag-Pb-Sb	Ag-Pb-Sb	Al-silicon eutectic
Base electrode	Mo, Ni-plated or Ta	Mo, Ni-plated	Mo, Ni-plated

and thick enough to ensure uniform wetting of parts which have dimensional tolerances. The solder is carefully cleaned before use.

The counterelectode solders for use with *p*-type silicon have the same requirements as the base electrode solders for use with *n*-type silicon except that traps may be present in the base solders provided they will not diffuse into the bulk silicon and lower the diffusion length below the counterelectrode-base extension.

3. *Base Solder*

The base solder to *n*-type material should have all the desirable mechanical properties noted in the discussion of the counterelectrode solder and should in addition produce an n^+ region in the silicon near the contact. Again, the same base materials of Au, Cu, and Ag are possible. Al is quite soluble in silicon and diffuses faster in silicon than the group V elements so that doping of Al with these elements will not produce *n*-type recrystallized silicon regions.

It is desirable in considering base contacts and necessary in producing a counter-electrode alloy to *p*-type material to dope the metals Au, Cu, and Ag with group V elements. In a vacuum fusion process, it is convenient to use Sb as the element. However, As has been employed with similar results. The alloys of Au, Cu and Ag with Sb and As are discussed in Table III.

The Au-Sb alloys are brittle, but make possible base or counterelectrode solders provided the processing temperatures do not exceed 350 C. The Cu alloys appear to have higher solidus and liquidus temperatures than the corresponding Ag alloys with no particular advantages. The Ag-Sb and As alloys become more difficult to roll and brittle as the Sb and As contents are increased. On the other hand, the solidus temperatures are not reduced appreciably by traces of the elements. The Sb alloy is preferred over the As one for vacuum fusion operations.

An alloy of Ag-5 wt % Sb was first employed as base solder in the development work. However, an appreciable number of cells cracked during processing and in operation as a result of thermal stresses in the silicon. Fig. 2 shows the appearance of a cell that has been destroyed by excessive thermal stresses. A very sensitive method of detecting stress in silicon cells is the measurement of forward drops at high currents as a function of number of temperature cycles (−55 C to 200 C) to which the cells have been subjected. If stresses are excessive, a gradual increase in forward drop will result. Reverse characteristics will behave in an erratic manner.

The high percentage, 5 by weight, of Sb in the Ag alloy need not be retained for doping considerations. However, another metal must be added to reduce the solidus temperature and the modulus of elasticity and increase ductility. Table IV lists some possibilities. Some of these cannot be used in a vacuum system. The most attractive are the alloys of Ag with germanium (Ge) or lead (Pb). Both alloys have been employed with success. However, the solidus and modulus of the Ag-Ge alloys are higher than in the corresponding Pb alloys. Hence, the Pb alloys are preferred for producing relatively strainfree silicon cells. Interesting ternaries of Ag, Pb, and Sb are listed in Table V. The solidus temperature can be tailored to the desired range.

The Ag-Pb-Sb solders have been employed as base solders on *n*-type materials. Cells made with such alloys are absolutely free of cracks as observed under microscopic examination. Hundreds of cells have been studied without observing failures.

The Au-Sb alloys have been employed as counterelectrode solders on *p*-type material. Al, Al-silicon eutectic, or the same with addition of Ga produce excellent base contacts to *p*-type bulk silicon.

The base solder thicknesses are again held to two or three mils. The range is such that uniform wetting can be obtained with the parts and jig tolerances permitted and excessive strain from the base electrode is avoided. Base solders are carefully cleaned before use.

4. *Counterelectrode*

The counterelectrode material must satisfy several requirements:

(*a*). It should have close to the same thermal expansion coefficient as the silicon over the entire range of temperatures encountered during operation and processing.

(*b*). It should have a bulk modulus high if the coefficient in (*a*) matches well and low if it does not.

(*c*). It should have high thermal conductivity.

(*d*). It should have fabrication possibilities and be reasonable in cost.

(*e*). It should be, if possible, inert to the action of etches employed in processing the basic diode.

(*f*). It should not contaminate the redeposited silicon.

(*g*). It should be wet by the counterelectrode solder in the range of fusion temperatures employed.

(*h*). It should be wet by solders used in attaching external leads.

Table VI lists the properties of a number of interesting materials, while Table VII gives values of thermal expansion coefficients measured on molybdenum (Mo) sheet, plain and clad, with a mil or so of metals. There appears to be no significant differences in expansion coefficients. The values averaged from 0 to −60 C were 5.2, 5.62, and 5.48 × 10^{-6} inch per inch C for plain Mo sheet, Mo sheet clad with Kovar alloy, and Mo clad with nickel (Ni) and stainless steel.

Chromium (Cr) cannot be obtained commercially in sheet form and should not be considered. While zirconium (Zi) has some interesting properties, the material is in short, government-controlled supply.

The iron (Fe)-Ni alloys do not match the coefficient of expansion of silicon over a sufficiently large range to be considered. Provided the counterelectrode and base solders do not freeze below the inflection point in the thermal coefficient of expansion, Kovar alloy may be used without cracking the silicon. It can be soldered with soft solders. However, the thermal conductivity is so low that it should not be considered a suitable electrode material except in applications of low power dissipation.

Of the elemental metals, tungsten (W) best matches silicon in coefficient of thermal expansion and has the highest thermal conductivity. The material is difficult to fabricate and relatively expensive. For very large silicon rectifiers, it appears to be the ideal base.

Tantalum (Ta) can be easily fabricated but is also expensive. It has a peculiar advantage in that it is relatively inert to the action of etches employed in the processing of cells. Thus, cells made with electrodes of Ta can be etched in batches without masking. The element can be welded, but cannot be wet by soft solders in encapsulation steps. It may be clad or plated with noble metals to permit soft soldering.

Mo can be easily fabricated and is less expensive than W or Ta. It is attacked by etches and must be protected in some manner. It is easily welded or soldered to with soft solders when clad with Kovar alloy, or plated with Ni or noble metals, tinned with Ag, tin (Sn), and indium (In). When the material is used as counterelectrode material, the surface next to the counterelectrode solder is not clad or plated.

5. *Bases*

The base electrodes are chosen on similar considerations to those previously discussed. Base electrodes may be clad, plated, or tinned on both sides without damage to the cell properties.

In practice, a variety of electrode materials have been employed, Ag-tinned Mo, Ni-plated Mo, Kovar- and Ni-clad Mo, Ta, platinum (Pt)-clad Ta, and W. For small diodes in which the component part expense has not been great, Ta has been employed because of its ease of fabrication and resistance to etch chemicals. In larger diodes, Ni-plated Mo has been found to be most economical and quite satisfactory.

FUSION OF CELL COMPONENTS

The parts as discussed under "Choice of Cell Components" are assembled in graphite jigs. In cases of the communication diode and larger power rectifier cells, the parts given in Table VIII are employed. Thickness and preparation are as described previously. Areas are chosen to handle the current desired.

The graphite jigs are heated rapidly in vacuum by radio-frequency coils to the fusion temperatures which approximate 900 C for the *n*-type and 800 C for the *p*-type cells. The jigs are then cooled rapidly to 500 C and more slowly to room temperature.

The base electrodes of the communication diode and the both electrodes of the power rectifier cells are masked, the former with Apiezon wax, the latter with Pb-Sn solder prior to etching. Etching takes place in a mixed acid etch. Careful washing, rinsing, and drying steps precede the encapsulation. Examples of finished rectifier cells are shown in Fig. 3.

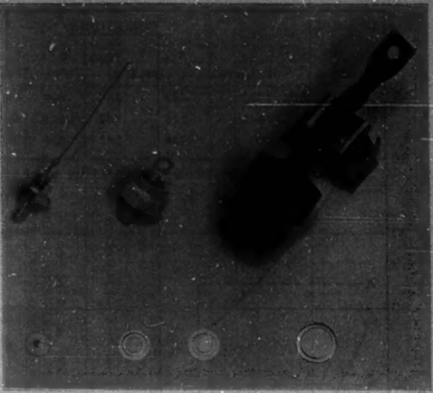


Fig. 3. Rectifier cells and encapsulation

Encapsulation and Heat Dissipation

The silicon rectifier cell must be protected from water vapor in order to obtain the optimum performance of the unit. The sensitivity to moisture is most pronounced in those units which most closely approach the nonequilibrium theory (small communication diodes) since in such units the external surface lifetimes are relatively high compared with those of poorer units and are more easily degraded. In addition, external connections must be made to the cell which will not introduce stresses in the silicon when the cell is subjected to all the environmental conditions imposed by industrial and military applications. Consequently, all silicon rectifiers are hermetically sealed in suitable capsules.

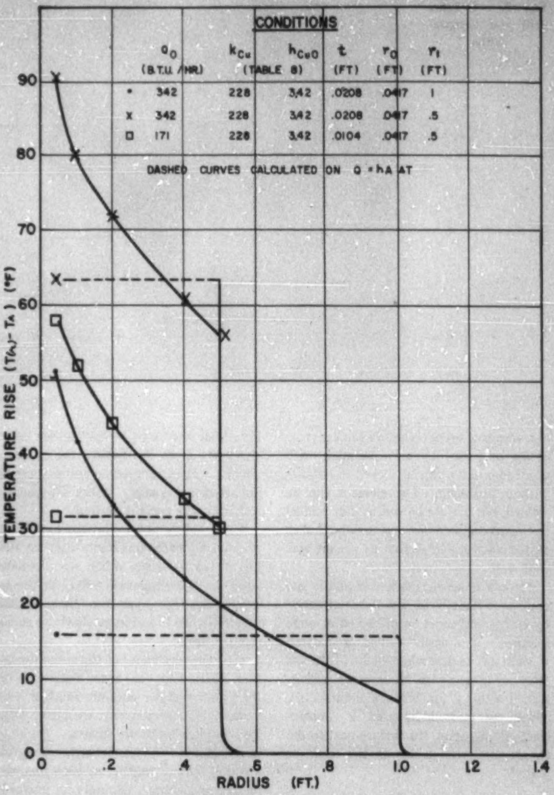


Fig. 4. Temperature drop in cooling fins by equation 8

In addition to supporting the external leads and giving humidity protection, the cases must provide for the dissipation of heat to a heat sink. In almost all applications, air cooling is desirable so that the heat sink is either the ambient atmosphere or a metal member to which the unit may be fastened in some manner. The latter must then dissipate the heat to the atmosphere. In this section, the general problem of dissipation of heat from the cells to the atmosphere will be considered although the rectifier units are available without fins, but with means of making connection to such.

Humidity Protection

In the course of development, various methods of protecting the cell against water vapor were used including, for example, coating with silicone rubbers and encasing in oxide power slurries, but all have been abandoned for general use in favor of hermetic sealing in gas. In selected applications, sealing in solid insulators may be of interest.

All cells are subjected to humidity and temperature cycling tests so that the requirements on glass or ceramic seals employed in the encapsulation are demanding. The case and seal assemblies are carefully cleaned and baked out in vacuum prior to use.

External Lead Connection

Because of the low ultimate strain of silicon, external lead connections must be made with considerable care. In order that free expansion might be retained in the x direction of Fig. 1, a flexible lead is connected to one electrode; in the cases under consideration, it is connected to the counterelectrode. In larger rectifier units, it is desirable to dissipate heat through both electrodes, consequently, a problem arises of making flexible parts of sufficiently low thermal resistance to permit significant heat dissipation. In smaller units, the problem reduces to one of making flexible contact with leads of sufficient cross section to handle the electric currents. A variety of flexible leads have been designed and used. In one case, a flexible lead was required which would permit the rectifier unit to operate in as small a space as possible under high centrifugal forces on the shaft of a generator. The flexible member required symmetry so that under starting and stopping conditions, the cell would not be subjected to dangerous forces. This member took the form of a flat notched hollow convex lens. In other cases, S-shaped members have been employed. There is a distinct advantage to the use of flat-dish-shaped or corrugated plates or bellows over braided copper wires in that the former may be more easily cleaned of flux employed in soft solder operations. In the soft soldering operations, solders melting safely above the cell operating temperature of 200 C must be used.

In the case of the communication diode shown in Fig. 3, the lead is an integral part of the counterelectrode. Flexible connection to this lead is easily made by welding. In the case of the small diode, requirements of very low leakages dictate extreme cleanliness and absence of flux in all operations.

Connection to the base electrode must be made with care to observe that the additional members, i.e., the solder and case metal, do not contribute to the total stress in the silicon. Equation 7 can be extended qualitatively to the 7-member case, however, symmetry is now lacking and bending certainly results. To dissipate heat adequately, the case base should be many times the thickness of the parts of the basic cell. The thermal coefficients of expansion of materials considered for the case will be high. Consequently, severe strains can result unless the soft solder plastically deforms. It has been found that lead solders are completely satisfactory. Tin may also be used as solder with smaller cells.

The technique employed in the fabrication of the communication diode provides the base of the case in the fusion operation. After this operation the header is welded to the base.

Vol.
Pt. 1
1956

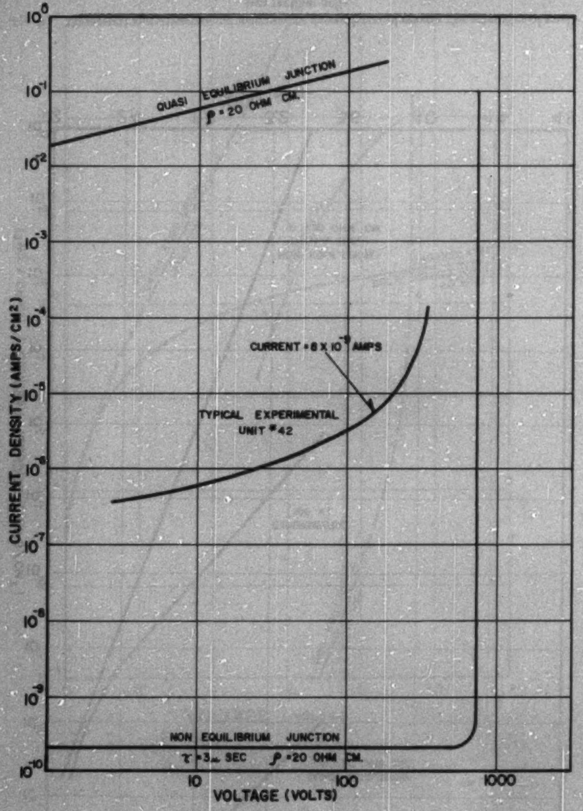
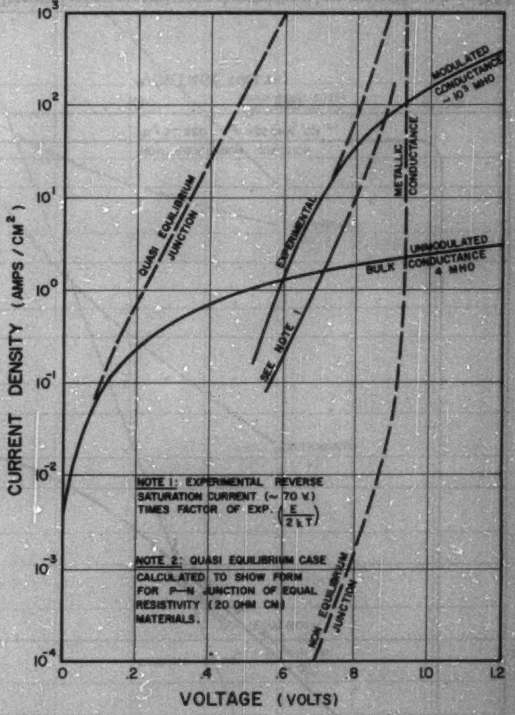


Fig. 5. Theoretical and experimental reverse characteristics of small-area rectifier cells, T = 298 K



NOTE 1: EXPERIMENTAL REVERSE SATURATION CURRENT (~ 70 V.) TIMES FACTOR OF EXP. $\left(\frac{E}{2kT}\right)$

NOTE 2: QUASI EQUILIBRIUM CASE CALCULATED TO SHOW FORM FOR P—N JUNCTION OF EQUAL RESISTIVITY (20 OHM CM) MATERIALS.

Fig. 6. Theoretical and experimental forward characteristics of small-area rectifier cells, T = 298 K

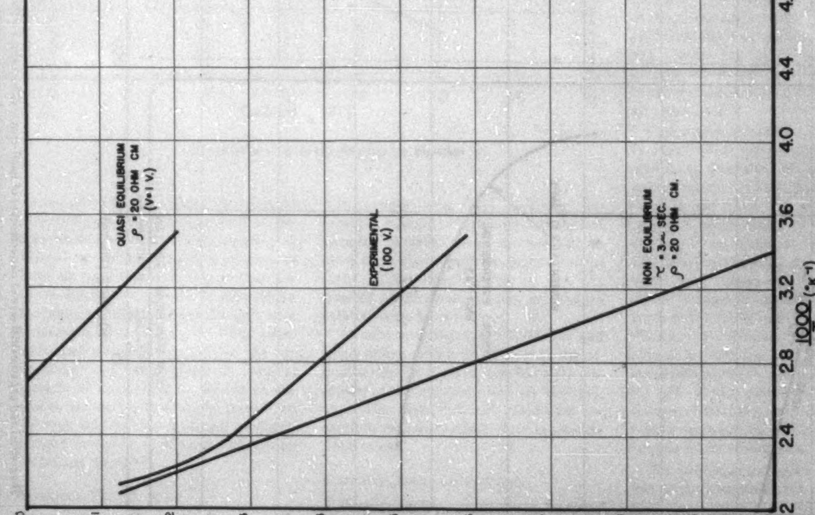


Fig. 7. Theoretical and experimental temperature dependence of reverse characteristics of small-area rectifier cells.

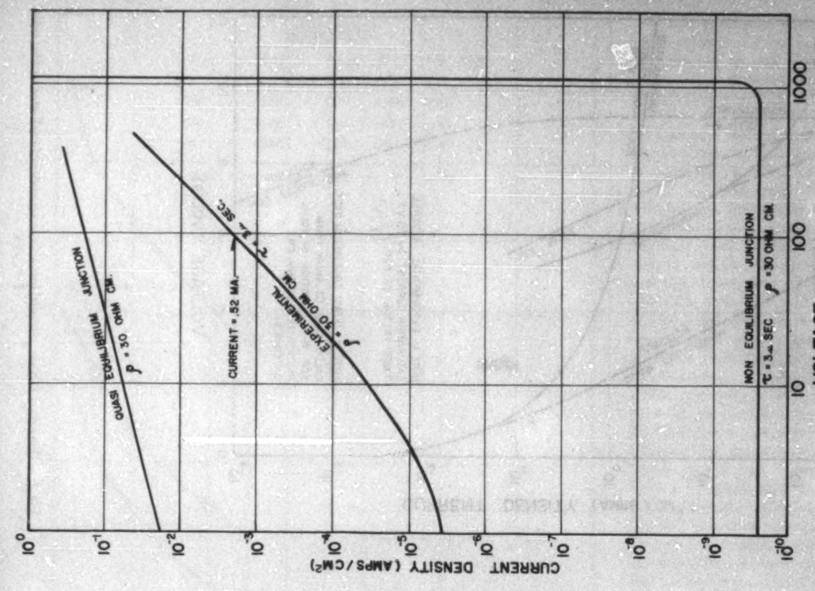


Fig. 8. Theoretical and experimental reverse characteristics of large-area rectifier cells, T = 298 K

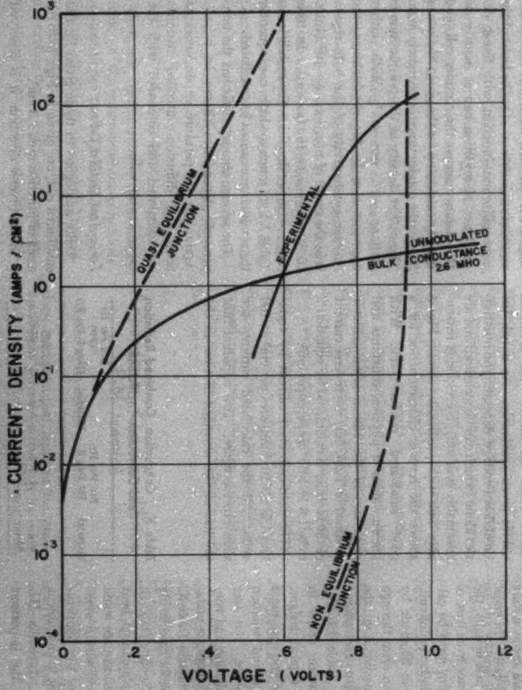


Fig. 9. Theoretical and experimental forward characteristics of large-area rectifier cells, T = 298 K

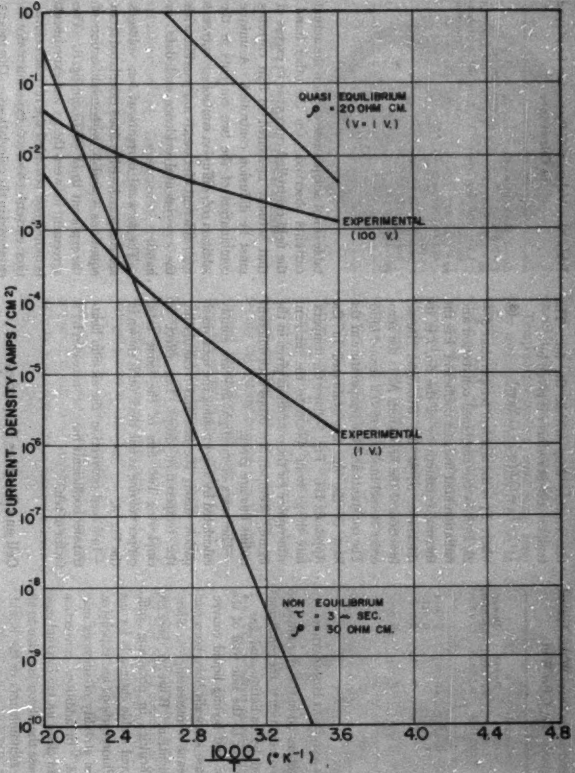


Fig. 10. Theoretical and experimental temperature dependence of large-area rectifier cells

Table IX. Temperature Drops Within Typical Encapsulations, C Per Watt

	Communications Diode, 0.002 Cm²	Power Rectifier, 0.3 Cm²	Power Rectifier, 1.26 Cm²
Base solder	0.42	0.006	0.0014
Base	8.65	0.115	0.0275
Soft solder		0.244	0.058
Spreading resistance	0.86	0.134	0.075

CASE DESIGN AND HEAT DISSIPATION

The heat generated in the case must be dissipated by conduction to the contiguous terminals and by radiation and gas or liquid convection to the side wall of the case. Liquid and boiling liquid convection to the case sidewalls have been considered as means of increasing the allowable heat dissipation. However, there are practical difficulties in obtaining sufficiently pure and stable chemicals for use at 200 C. Fluorocarbons are expensive. The long-term stability of silicones may be questioned. In addition, convection liquids tend to clean up the internal case parts and expose the silicon junction to all the resulting contaminants in the liquid.

The radiation from cell to case may be neglected if the case is properly designed since the temperature of the case will approach that of the cell.

The conduction heat losses through one or both of the Mo terminals of the cell can be readily calculated. Table IX lists representative temperature drops within elements of the cell and case. In most common and convenient case designs, the thermal spreading resistance from base of the case to the roots of fins must be calculated. An exact calculation is difficult, but an approximation can be made by averaging the cross-section areas through which the heat flows passing from the bottom of the case to the fins in estimating the thermal resistance. Table IX also lists estimated temperature drops in spreading resistances for typical designs.

The heat at the fins of the case must be dissipated by convection and radiation to the ambient atmosphere. For free convection, combined radiation and convection coefficients of different case materials are estimated to be those in Table X. The usual simple formulas for calculating temperature drops across cooling plates in terms of total areas of plates and heat transfer coefficients are inadequate when considering the case of cylindrical symmetry. An expression for the temperature drop from a finned case to an ambient gas has been developed, Appendix III, which does not neglect the important temperature drop within the fin. The temperature at radius r is given by

$$T(r)-T_A=C_1J_0(qr)+C_2N_0(qr) \qquad (8)$$

where T_A is the ambient temperature $q^2=2h/kt$, h is the combined convection and radiation heat transfer coefficient, k is the thermal conductivity of the fin, t is its thickness, J_0 is the zero-order Bessel function of the first kind, N_0 is the zero-order Bessel function of the second kind. The constants and the temperature at the base of the fin are evaluated in the Appendix III. Fig. 4 shows the temperature drops from silicon to the ambient atmosphere for the conditions given in the figure. Temperature drops with different radius fins are given.

The cases of multiple fins are simply calculated by apportioning the total dissipation among the fins and recalculating the temperature drop. The effect of employing two fins of the same total copper volume as one thick fin is shown in the Fig. 4.

In forced convection, the air-film heat-transfer coefficients may be increased by a factor of ten.

Cell and Rectifier Properties

PROPERTIES OF THE BASIC CELLS

Small-Area Rectifier Cells

The breakdown voltage of small-area rectifiers (Fig. 3, junction area 0.0020 cm²) shows some agreement with predicted values given previously. As noted, the breakdown voltage, depending on resistivities alone (which can be measured and are reasonably uniform), corresponds reasonably well with predicted values. Fig. 5 shows the reverse characteristics for a good small-area rectifier cell. With known values of physical constants and properties of the silicon semiconductor assumed in Table XI, the reverse current of the rectifier is calculated by equation 2 to be 1.8×10^{-10} amperes per cm².

The field-controlled current by equation 1 is 2×10^{-1} amperes per cm² at 100 volts. All the calculated currents are shown in the figure. It is apparent that the use of the averaged lifetime in the table is not adequate to explain the actual currents observed. On the other hand, the field-controlled current is in excess of that actually measured so that currents must be diffusion controlled. A unique combination of the two currents or the addition of diffusion-controlled currents with different lifetimes is not possible with the experimental evidence and data on hand.

Table X. Conduction, Combined Radiation and Convection Coefficients

Material	Btu Per Hr Ft² Deg F	Material	Btu Per Hr (Ft²) (Deg F Per Ft)
h_{CuO}	3.42	k_{Cu}	228
h_{Cu}	1.52	k_{Al}	89.6
h_{Al}	1.38		

Table XI. Values of Si Parameters Employed in Calculations

$E_g=1.21$ electron volts
$T=298$ K
$\mu_n=5.5\times10^6T^{-1.5}\sim 1{,}100$ cm² per volt-second
$\mu_p=2.4\times10^8T^{-2.3}\sim 490$ cm² per volt-second
t (semiconductor thickness) $=1.27\times10^{-2}$ cm
$p_p\sim10^{18}$ per cm³
$\tau_p=3\mu$ sec
$\rho_n=20$ ohm-cm
$A=1.5\times10^{33}$ cm⁶/°K³
V_D (intrinsic to 20 ohm-cm) $=0.321$ electron volts

The forward currents at low voltages are compared with values calculated from equations 4 and 5, with modification of the exponent to $(V/2kT)$, in Fig. 6. The agreement between theory and experiment is quite good provided the reverse saturation current observed experimentally is employed in the calculations. This means that the term exp $(V/2kT)$ is essentially correct in the range of voltage of 0.5–0.6 volts although the diffusion length, or lifetime, is not adequately evaluated. The asymptotic value of the slope of the characteristic is compared with the value calculated on the basis of the quasi-equilibrium junction. On the basis of the nonequilibrium junction with linear dependence of injection carrier density on the distance across the rectifier cell, and density calculated for the given lifetime of 3 μsec, the conductance is extremely high. The carrier density at the *p-n* junction is perhaps 10^{20} per cm³. With either approximation to the conductance, the average conductivity of the bulk region is in the range of $n(\mu_n+\mu_p)\ q\sim10^{20}\ (1.6\times10^{-19})\ (\mu_n+\mu_p)=16\ (\mu_n+\mu_p)$. In this range of carrier densities, the mobilities in Table XI are much too high, consequently the conductance cannot be calculated.

The reverse characteristic has the temperature dependence shown in Fig. 7. Again, the temperature dependences of field- and diffusion-controlled currents are calculated and compared with the experimental values.

Large-Area Rectifier Cells

Data, analogous to those given for the small-area cell, are given in Figs. 8–10 for a cell of junction area 0.3 cm².

In the reverse direction, it is observed

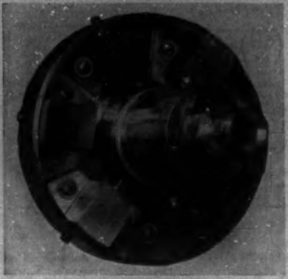


Fig. 11. Rotating rectifier assembly

that the current does not show a saturation value. The field-controlled current, on the other hand, is much larger than actually observed. Consequently, the lifetime is considered to vary across the junction, particularly at or near the surface edges. A unique description of the current is not possible with the data on hand.

The forward characteristic agrees much better with the values predicted by theory. The conductance of the carrier-modulated bulk is lower than calculated, which indicates room for improvement in the forward drop.

The temperature dependence of the leakage of cells shown in the figure is representative of that of many production cells. It should be noted that the current increases less with temperature at high voltages than at low voltages. In many cases at sufficiently high voltage, the leakage currents decrease with temperature. If an effective lifetime for the junction is calculated such that the theoretical reverse current at high temperature (which current is reasonably constant with voltage) matches that observed experimentally, an exponential dependence of lifetime on temperature results on matching the theoretical saturation current with the experimental at an arbitrary low voltage, say 1 volt. It is apparent that the power rectifier cells presented in this paper cannot be adequately described in terms of lifetime parameters.

PROPERTIES OF RECTIFIERS

Complete rectifiers have been tested under all the environmental conditions required under the *MIL-E-5272-A* specification. Such tests have included humidity, salt-spray, vibration, and temperature-cycling tests. With good encapsulation and solution of cracked header problems, no difficulty was expected or experienced in salt-spray and humidity tests.

Temperature-cycling tests were carefully conducted to detect any increase in forward drops indicative of cracking silicon. With the solders described in the previous section, success has been achieved in eliminating cracking as an important problem in the silicon rectifier.

Ageing tests have been in progress for over 15,000 hours. The characteristics of the cells have been stable.

Rectifier units have been mounted in various assemblies to test the performance under different load conditions in free and forced air convection. The rectifier stack shown in Fig. 11 was designed to rectify the exciter current in a brushless aircraft generator. In such an application, the rectifiers were cooled by oil at 150 C flowing through the shaft of the generator. The generator was subjected to overspeed tests to 13,000 rpm so that the cells were required to withstand centrifugal forces of 10,000 g.

The single-phase assembly in Fig. 12 was designed for industrial applications requiring about 30–40-ampere d-c output at voltages to 200 volts in free air convection.

A high-power application required the use of the base cell in Fig. 3. This single rectifier cell, properly mounted, is capable of handling currents as high as 1,000 amperes on short-circuit condition for about 5 seconds, and is rated for average current of 150 amperes.

The foregoing illustrate a few of the applications of silicon rectifiers. A number of difficulties have arisen in applying silicon rectifiers. One of the more serious of these is the correct rating of the rectifiers for circuits in which current surges are obtained. Other problems have arisen in the aging of parts of the rectifiers. Such problems, as they have arisen, have been solved by careful process control and rating. It is expected that further difficulties will arise as the silicon rectifier is given wider and wider application. However, the physical and chemical stability of the correctly constructed and carefully encapsulated basic cell are such that the general successful application of the silicon rectifier may be looked for within the next few years.

Appendix I. Quasi-Equilibrium P-N Junction

Assuming an abrupt *p-n* junction of two semiconductors of equal carrier densities, these independent of position, the quasi-equilibrium reverse current can be calculated realizing that the total current across the junction is continuous. Schottky[8] has shown that the current through a metal-semiconductor junction with internal contact difference of potential of V_D is given at reverse voltages over about 0.1 volt by

$$i=\sigma_R\left(\frac{8\pi q n_A^-}{\epsilon}\right)^{1/2}(V_D+V)$$

(electrostatic units)

where, in electrostatic units, i is the current per square centimeter, σ_R is the conductivity at the junction, n_A^- is the density of ionized acceptors in the vicinity of the junction, ϵ is the dielectric constant, V is the applied voltage.

Realizing that the foregoing is the current that flows when V is the drop in voltage in the material on one side of the *p-n* junction, provided σ_R is now the conductivity at the *p-n* junction, V_D is the contact difference of potential intrinsic to material on one side of the junction, the terms can be subscripted one and two for the two sides of the junction and the voltage drops summed to give the following

$$i=\sigma_R\frac{\left(\frac{8\pi q n_A^-}{\epsilon}\right)^{1/2}(V_{DT}+V)^{1/2}}{\sqrt{2}}$$

(electrostatic units)

where V_{DT} is the total of contact potentials V_{D_1} and V_{D_2}. This may be rewritten as in equation 1.

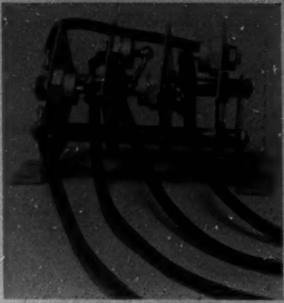


Fig. 12. Medium-power industrial bridge

Appendix II. Stress in a Silicon Junction

In the case of the 5-element sandwich discussed in this paper, a rather symmetrical situation exists. Consequently, a first approximation to the stress in the silicon can be obtained by assuming that bending is absent, and all strain occurs in the plane of the flat elements.

In cylindrical co-ordinates, the radial strain $\delta R/R$ is the same as the strain in the θ_1 direction.

Consider a 5-element sandwich in which elements 3, 4, and 5 are rigidly connected at temperature T_4 by freezing of element 4, elements 1, 2, 3, 4, and 5 are joined by freezing of element 2 at T_2 and stress in element 3 is to be calculated at temperature T. A system of equations:

$$\frac{S_n}{E_n}=\frac{R(T)-R(T_2)[1+\alpha_n(T-T_2)]}{R}\qquad n=1,2$$

$$\frac{S_n}{E_n}=\frac{R(T)-R(T_4)[1+\alpha_n(T-T_4)]}{R}\qquad n=3,4,5$$

$$\sum_n S_n t_n=0\qquad n=1\text{–}5$$

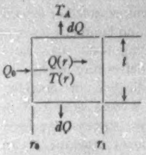
$$\frac{S_n{}^1}{E_n}=\frac{R(T_2)-R(T_4)[1+\alpha_n(T_2-T_4)]}{R}\qquad n=3,4,5$$

$$\sum_n S_n{}^1 t_n=0\qquad n=3,4,5$$

can be solved for the stress S_3 at T in equation 7.

Appendix III. Temperature Drops in Cylindrical Fins

Consider the flow of heat through a cylindrical cooling fin shown in the following cross section:



$$Q(r)=2\pi rkt-\frac{dT(r)}{dr}$$

$$dQ=-h(T(r)-T_A)4\pi r dr$$

Hence

$$\frac{d^2T(r)}{dr^2}+\frac{1}{r}\frac{dT(r)}{dr}+\frac{2h}{kt}(T(r)-T_A)=0$$

or with $T^1=(T(r)-T_A)$, $q^2=\frac{2h}{kt}$,

$$\frac{d^2T^1(r)}{dr^2}+\frac{1}{r}\frac{dT^1(r)}{dr}+q^2T^1=0$$

The solution of this equation is the zero order Bessel function

$$T^1=C_1J_0(qr)+C_2N_0(qr)$$

The boundary conditions, at $r=r_0$,

$$T_0{}^1=(T_0-T_A)$$

$$\left(\frac{dT^1}{dr}\right)_0=\frac{-Q_0}{2\pi ktr_0}$$

allow evaluation of the constants in terms of T_0. Evaluation in terms of Q_0 is possible by realizing that

$$Q_0=-\int dQ=h\int_{r=r_0}^{r=r_1}(T(r)-T_A)4\pi r dr$$

and substituting for $T(r)$ from the equation 8.

References

1. THE THEORY OF CRYSTAL RECTIFIERS, N. F. Mott. *Proceedings*, Royal Society, London, England, vol. 171, 1939, p. 27.

2. ZUR QUANTITATIVEN DÜRCHFUHRUNG DER RAUMLADUNGS UND RANDSCHICHTTHEORIE DER KRISTALLGLEICHRICHTER, W. Schottky, E. Spenke. *Wissenschaftliche Veroffentlichungen aus dem Siemens-Werken*, Berlin-Siemensstadt, Germany, vol. 18, 1939, p. 225.

3. ON THE CONTACT RESISTANCE OF SEMICONDUCTORS, B. Davydov. *Journal of Physics*, Moscow, USSR, no. 1, 1939, p. 167.

4. THE THEORY OF P-N JUNCTIONS IN SEMICONDUCTORS AND P-N JUNCTION TRANSISTORS, W. Shockley. *Bell System Technical Journal*, New York, N. Y., vol. 28, 1949, pp. 435–89.

Application of Junction Transistors to Carrier-Frequency Computing Amplifiers

W. A. CURTIN
ASSOCIATE MEMBER AIEE

THE PURPOSE of this paper is to describe some of the considerations involved in designing a transistorized summing amplifier which could be used as a building block in a carrier-frequency analogue computer. The basic design philosophy was necessitated by the quality of the transistors commercially available at the start of the development work in mid-1954. Typical transistors are the early type *2N34* and the later type *2N109*.

A need for transistorized summing amplifiers exists in the a-c computers of military fire-control systems where long life, miniature size, and low power consumption are essential requisites. Such a computer may be subjected to extreme environmental conditions, have a limited power source, and may be required to function with little or no maintenance. The transistor, because of its ruggedness small size, and low power requirements, would make a subminiature computer possible. Thus, it appears that the transistor is ideally suited to such an application.

With these thoughts in mind, it is felt that if transistorized components, such as a summing amplifier, could be developed to perform as well as a high-quality vacuum-tube device, then the goals of long life, subminiaturization, and power conservation would be partially realized.

Background

Negative feedback is usually employed to maintain the high accuracies required of a summing amplifier for computer applications. The general form of such an amplifier is shown in Fig. 1. Before proceeding with those considerations peculiar to transistors, it might be well to review briefly part of the theory of the circuit shown in Fig. 1, limiting the discussion to amplifiers having only resistive feedback elements, and the main concern being dynamic accuracy and stability.

For the circuit of Fig. 1, the output voltage is given by

$$E_o=\left[\frac{R_F}{R_{i1}}E_{i1}+\frac{R_F}{R_{i2}}E_{i2}+\ldots\frac{R_F}{R_{ij}}E_{ij}\right]\frac{A\beta}{1+A\beta}\tag{1}$$

where

$A\beta$ = open-loop gain

For reasons of stability, $A\beta$ is necessarily finite; hence the closed-loop gain is not exactly equal to the indicated resist-

---

Paper 56-758, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted November 15, 1955; made available for printing June 1, 1956.

W. A. CURTIN is with the Radio Corporation of America, Moorestown, N. J.

This paper is abstracted from a master's thesis submitted to the Department of Electrical Engineering of Massachusetts Institute of Technology, Cambridge, Mass., August 1955. The author is indebted to Prof. T. F. Jones, Jr., for his interest, encouragement, and suggestions as thesis supervisor.

Vol. 7
Pt. 1
1956

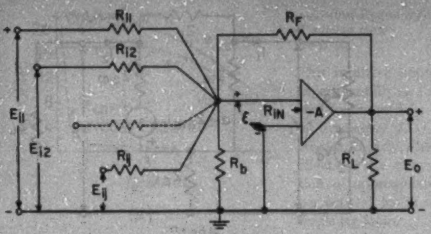


Fig. 1. General form of linear parallel feedback amplifier; see equation 1

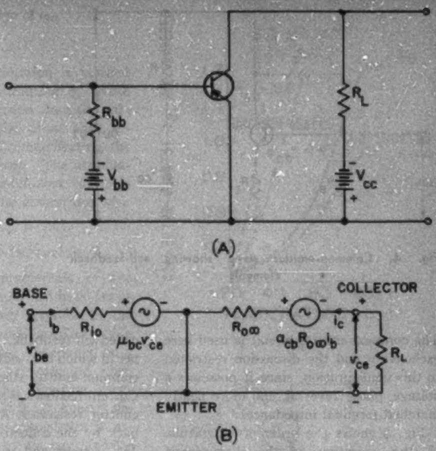


Fig. 2 (right). Common-emitter amplifier

A—Schematic representation
B—Linear incremental equivalent circuit

ance ratio of equation 1, and therefore gives rise to a static error which may be compensated for by altering the value of either R_i or R_f. It can be shown that a finite amplifier input resistance R_{in}, or input shunting resistance R_b does not affect the closed-loop gain, but only affects the gain stability factor $A\beta$.

More important is the error resulting from variations of $A\beta$ from a design center value caused by transistor (or tube) aging, component tolerances, environmental conditions, and so on. Here the question arises as to how much feedback is required to satisfy a given error specification. The answer to this problem lies in how much $A\beta$ will change with respect to the design center value, $(A\beta)o$. If the maximum expected $A\beta$ variation is expressed as a fraction of $(A\beta)o$, then the value of $(A\beta)o$ required to hold the error to less than a specified amount may be calculated from equation 2

$$(A\beta)o=\left(\frac{1}{e_d}-1\right)\left[\frac{1-(A\beta)n}{(A\beta)n}\right]-1 \qquad (2)$$

where

$(A\beta)n=A\beta/(A\beta)o$
e_d = error

A useful set of design curves is obtained if $(A\beta)o$ is plotted against $(A\beta)n$ with e_d as a parameter. For example, in order to insure a maximum error of 0.1% (per cent), $(A\beta)o$ need only be 110 if $A\beta$ can decrease by 10%, but must be approximately 1,000 if $A\beta$ can decrease by 50%.

From the point of view of stability, a large amount of feedback is troublesome. Alternatively, the problem of stability is significantly reduced if $A\beta$ can be constrained to vary between narrow limits, which, as seen from the preceding, will also permit the realization of a low error. It will be shown later that the inherent limiting phase characteristic of the transistors available for this investigation sets a rather restricted upper limit of stable loop gain which may be realized.

In addition to these notions, all the other well-known virtues of negative feedback apply to a transistor feedback amplifier.[1] Since the design techniques of linear vacuum-tube feedback amplifiers are well known,[2] one's first tendency is to determine whether they might be applicable to an equivalent transistor amplifier. With recognition of the transistor as a new circuit element, a determination should be made of its frequency response, and of the behavior of its gain and terminal impedances, as would normally be done in the designing of the vacuum-tube amplifier.

Gain and Terminal Impedance Variations

It is well known that the intrinsic transistor parameters, which determine the input and output impedances and gain characteristics, vary rather widely between units as a result of the manufacturing process. Within a given unit, variations exist and are attributable to bias conditions and temperature.[3]

For the purposes of the present discussion it is more convenient to consider the variation of the parameters of the low-frequency equivalent circuit rather than that of the intrinsic resistance elements and amplification factors. A convenient form of the low-frequency common-emitter equivalent circuit is shown in Fig. 2 (above), where, if $R_{o\infty} >> R_L$.

R_{io} = short-circuit input resistance $= r_b+(1+\alpha_{cb})r_e$
$R_{o\infty}$ = open-circuit output resistance $= r_c/(1+\alpha_{cb})$
μ_{bc} = reverse open-circuit amplification factor $\approx r_c/\left(r_e+\dfrac{r_c}{1+\alpha_{cb}}\right)$
α_{cb} = short-circuit collector-base current gain $\approx \alpha/1-\alpha$

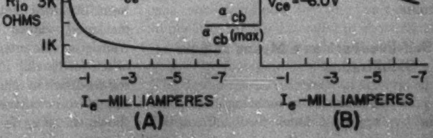


Fig. 3

A—R_{io} versus emitter current
B—α_{cb} versus emitter current

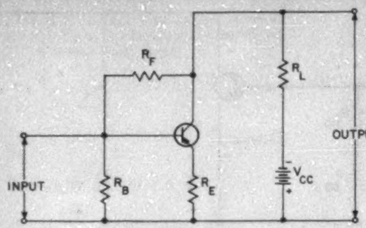


Fig. 4. Common-emitter stage showing self-feedback elements

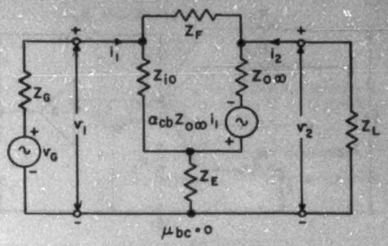


Fig. 5. Common-emitter equivalent circuit for computing the equations of Table I

The common-emitter circuit is used here exclusively, and the discussion restricted to this configuration, since it possesses a voltage phase reversal, and more nearly matched terminal impedances.[4]

Fig. 3 shows the order of magnitude of the variation of the short-circuit input resistance R_{io} and short-circuit current gain α_{cb} as a function of the emitter bias current. Since stage gain is proportional to the load resistance, it is clear that for cascaded common-emitter stages, the driving-stage gain will vary directly as the input resistance of the driven stage varies. Such variations may come about when one attempts to design a circuit, based on manufacturers' typical data, which may differ materially from that of the transistors at hand. As a consequence of this phenomenon alone, the net gain change per stage may be of the order of 6 db (decibels), which is further compounded in the case of multistage amplifiers. In circuits where the operating point is critical, such variations are intolerable, and yet one does not wish to be faced with the alternative of selecting transistors. Accordingly, a means of reducing such variations is sought in the hope of making transistors with vastly different characteristics look more alike.

The factor of temperature tends to compound the circuit designer's problems. The major effect of a temperature rise is an exponential increase of the collector leakage current I_{co}, causing a corresponding increase of the collector current I_c. This results in a change of the quiescent operating point and, as far as a-c amplifiers are concerned, is equivalent to the effect caused by α_{cb} variations between units.

Self-Feedback as a Means of Gain Stabilization

An effective way of reducing variations is by applying local feedback, aptly called self-feedback. The particular manner in which the feedback is applied to the common-emitter stage is shown in Fig. 4 Current feedback is obtained by the series emitter resistance R_e, and voltage feedback by the collector-base resistance R_f. The current and voltage gains are stabilized by the feedback action, while the terminal impedances are magnified in magnitude but not stabilized. Stability of the magnitude of the terminal impedances results from the base shunting resistance R_b and the collector load resistance R_L, which are, in general, chosen to be an order of magnitude smaller than the value of the terminal impedances in the presence of self-feedback. All of the external resistance elements in the circuit of Fig. 5 contribute toward stabilization of the operating point.

Analysis of the circuit of Fig. 5 yields expressions for the gains and terminal impedances in the presence of self-feedback. The results, given in Table I for the common-emitter amplifier, are extremely useful in predicting transistor performance using typical manufacturer's data. The data of Fig. 6 illustrate some of the virtues of self-feedback. It is seen that, with current feedback alone, the short-circuit input resistance R_{io} is no longer a violent function of emitter current, and, further, that the magnitude is magnified but not stabilized for transistors having different values of α_{cb}. Fig. 6 also shows that R_{io} increases almost linearly with R_e, and in the linear region is given approximately by $\alpha_{cb}R_e$. The last graph of Fig. 6 shows that the voltage gain is almost independent of the value of α_{cb} as R_e is increased. Excellent agreement has been obtained between calculated and measured values of these and other transistor characteristics in the presence of self-feedback.[5] The beneficial effect of self-feedback in reducing the sensitivity of the collector current to changes in I_{co} is shown in Fig. 7, where the operable temperature range is extended considerably over the uncompensated case. It may be shown that for the circuit of Fig. 7 the change in I_c, caused by a change of α_{cb}, is reduced by a factor of 15.[6]

Frequency Characteristics

Those frequency characteristics of the common-emitter amplifier which are pertinent to the stability problem will now be examined.

The quantity of interest is the voltage gain which is given by equation 3 in which K_v is the voltage gain from the base-input terminals (excluding source impedance) to the output.

$$K_v = -\frac{K_i}{Z_i} Z_L \tag{3}$$

where

K_i = current gain
Z_i = input impedance
Z_L = load impedance

If, as is the case for cascaded common-emitter stages, the condition of equation 4 is satisfied, then K_i and Z_i of equation 3 may be replaced by their short-circuit values, namely, α_{cb} and Z_{io} respectively.

$$Z_{o\infty} >> Z_L \tag{4}$$

where

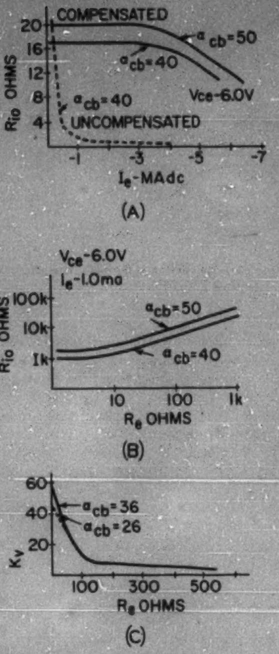
$Z_{o\infty}$ = open-circuit output impedance

Since both current gain and input impedance of a junction transistor are functions of frequency, it is more convenient to consider the frequency characteristics of their ratio, namely, the short-circuit collector-base transconductance defined as:

$$g_b = \frac{\alpha_{cb}}{Z_{io}} \tag{5}$$

Thus under the condition of equation 4 the voltage gain may be expressed by equation 6.

$$K_v = -g_{cb}Z_L \tag{6}$$



(A)

(B)

(C)

Fig. 6. A—R_{io} versus emitter current I_e
B—R_{io} versus external emitter resistance R_e
C—Voltage gain K_v versus external emitter resistance R_e

SHORT-CIRCUIT CURRENT GAIN α_{cb}

Charge carriers (holes for p-n-p transistors) injected into the emitter reach the collector by a process of diffusion through the base region, and so require a small but finite time to reach the collector. Hence a delay should be expected, but, as the frequency is increased, all the charge carriers do not arrive at the collector simultaneously, with the result that, if the time of arrival amounts to an appreciable portion of a cycle, some of the carriers tend to cancel the effect of others, and the collector current begins to fall off with increasing frequency.[7] Hence, to a first approximation, one should expect the current transfer function to be of the form given in equation 7.[8]

$$\alpha_{cb}(j\omega) = \alpha_{cbo}\frac{e^{-j\omega T}}{1+j\Omega} \tag{7}$$

where

α_{cbo} = low-frequency value of α_{cb}
T = delay time
$\Omega = \omega/\omega_{cb}$

This intuitive approximation is verified experimentally as shown in Fig. 8. Note that for frequencies greater than the cutoff frequency (ω_{cb}), the phase shift increases linearly, thereby manifesting the effects of the delay term. The attenuation of approximately-6db/oct (decibels per octave) justifies the assumption of a single lag factor.

SHORT-CIRCUIT INPUT IMPEDANCE Z_{io}

The frequency characteristic of the input impedance is important in so far as it affects the frequency response of the voltage gain. The loading effects of a frequency variant input impedance on a preceding stage may be neglected when self-feedback and a base shunting resistance are employed. If the frequency dependence of the input impedance is assumed to be caused by an emitter-base diffusion capacitance,[9] then it can be shown that the input impedance exhibits low-pass filter characteristics as given by equation 8 and shown in Fig. 9.

$$Z_{io}(j\omega) = R_{io}\frac{1+j\frac{1}{\eta}\Omega}{1+j\Omega} \tag{8}$$

where

R_{io} = low-frequency value of the input impedance
η = frequency- or ultimate-attenuation ratio $= R_{io}/(r_b+r_e)$ for the common emitter stage

Excellent agreement is obtained between calculated and measured values of the frequency characteristics of the input impedance.

SHORT-CIRCUIT TRANSCONDUCTANCE g_{cb}

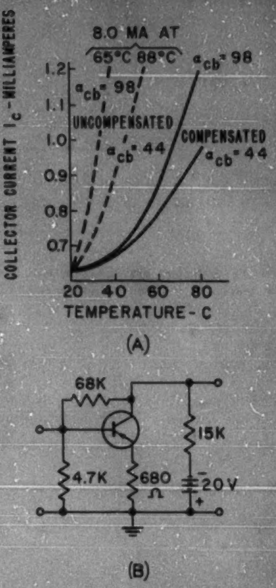
From equation 5 and a knowledge of the frequency characteristics of α_{cb} and Z_{io}, g_{cb} may be expressed as

$$g_{cb}(j\omega) = g_{cbo}\frac{e^{-j\omega T}}{1+j\frac{1}{\eta}\Omega} \tag{9}$$

where

g_{cbo} = low-frequency value of g_{cb}

Equation 9 indicates that the magnitude of g_{cb} has a "break frequency" at $\eta\omega_{cb}$ and then decreases at the rate of −6db/oct. Measured characteristics of g_{cb} agree well with equation 9. Fig. 10 illustrates the graphical construction of the g_{cb} characteristic from the α_{cb} and Z_{io} characteristics. This is obtained by subtracting decibel magnitudes and phase angles as required by equation 5. At the g_{cb} break frequency the phase angle is of the order of −150 deg (degrees) rather than the characteristic −45 deg of a pure lag network. This means that when transistor feedback amplifiers are designed, particular attention should be given to the phase angle of g_{cb} which is quite different from what is normally encountered in vacuum-tube amplifier design in the same frequency range. Self-feedback has little beneficial effect on the rapidly changing phase angle at high frequencies. Consequently, the maximum stable loop gain which can be realized will be ultimately limited by the phase characteristic of g_{cb}, and not by the phase shift from stray circuit capacitance, etc., (as in the case of a vacuum-tube amplifier).



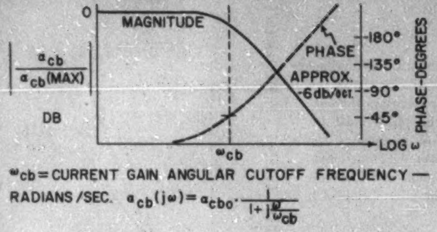
(A)

(B)

Fig. 7. A—Compensated and uncompensated collector current versus temperature
B—Circuit used for the data of A

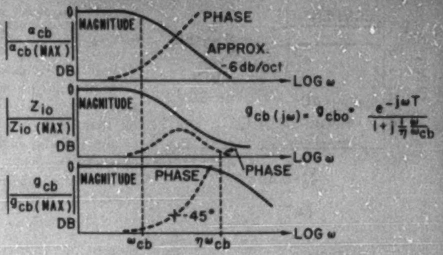
Computing Amplifier Design

The amplifier described in the following was not developed for any particular application, but is intended to show that contemporary commercial transistors are usable in high-gain feedback amplifiers. The design specifications given are typical of those imposed on a summing amplifier for use in a fire-control computer. In



ω_{cb} = CURRENT GAIN ANGULAR CUTOFF FREQUENCY — RADIANS/SEC. $\alpha_{cb}(j\omega) = \alpha_{cbo} \cdot \frac{1}{1+j\frac{\omega}{\omega_{cb}}}$

Fig. 8. Magnitude and phase characteristics of the short-circuit current gain α_{cb}, ω_{cb} = current-gain angular-cutoff frequency, radians per sec



$$g_{cb}(j\omega) = g_{cbo} \cdot \frac{e^{-j\omega T}}{1+j\frac{1}{\eta}\frac{\omega}{\omega_{cb}}}$$

Fig. 10. Illustrating the graphical construction of the $g_{cb}(\omega)$ characteristic from the $\alpha_{cb}(\omega)$ and $Z_{io}(\omega)$ characteristics

addition, the amplifier should be subminiaturized, thus precluding the use of high-Q inductance capacitance stabilization networks. For the present design a further, though perhaps unnecessary, restriction will not permit the use of transformers.

SPECIFICATIONS

Carrier frequency, 400 cycles per second ± 5%

Bandwidth, 100 cycles

Inputs, 2, independent

Input impedance, 100 kilohms resistive

Closed-loop gain, 1.0

Maximum dynamic error, ±0.1%

Minimum load impedance, 50 kilohms resistive

Operable temperature range, 20–65 degrees centigrade

PRELIMINARY DESIGN CONSIDERATIONS

For a closed-loop input resistance of 100 kilohms and for unity gain for each input, from equation 1 it is clear that $R_{i1} = R_{i2} = R_f =$ 100 kilohms. In the presence of self-feedback, it will be assumed that the total amplifier gain may decrease by 50% i.e., $(A\beta)n = 0.5$ which is approximately a 13% gain decrease per stage for a 5-stage amplifier. From equation 2 the required $(A\beta)o$ is 1,000 for an error of 0.1%. Since the design procedure is not unique, β may be arbitrarily chosen. However, the value chosen should not impose a severe amplifier-gain requirement. For this amplifier, β was chosen as 0.125, and consequently, as shown in Fig. 11, values of β, R_i, and R_f determine R_b'. To reduce the sensitivity of β to changes of the amplifier input resistance, and to reduce β frequency dependence, R_b should be small compared to R_{in}. In Fig. 11, where R_{in} = amplifier input resistance and R_b = base-shunting resistance, let

$$R_b' = \frac{R_b R_{in}}{R_b + R_{in}}$$

$$\beta = \frac{R_i'}{R_i' + R_f}$$

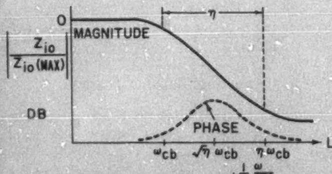
where $R_i' = R_{i1}$ in parallel with R_{i2} in parallel with R_b'. For $R_{i1} = R_{i2} = R_f = 100$ kilohms and $\beta = 0.125$, then $R_b' = 20$ kilohms. If $R_b = 1/3\ R_{in}$ then $R_b' = 3R_b/4 = 20$ kilohms, therefore $R_b \approx 27$ kilohms and $R_{in} \approx 81$ kilohms. Thus, the self-feedback resistances of the input stage are chosen to satisfy the input resistance requirement imposed by choice of a β value.

If the feedback resistance is very much larger than the open-loop output resistance and load resistance in parallel, and if direct transmission is neglected, then A is essentially equal to the amplifier voltage gain. For $A\beta = 1{,}000$ and $\beta = 0.125$, the required amplifier gain at the carrier frequency is 8,000. However, if it is assumed that the interstage compensation networks introduce about 2-db loss at the carrier frequency, then the amplifier gain must be increased to approximately 10,450 in order to account for this loss.

As a result of the input resistance requirement, the input stage will have a gain of approximately unity, and since an odd number of stages is required, the last two stages of a 3-stage amplifier must supply the total gain of 10,450, or a gain of approximately 102 per stage. This is not easily realizable with cascaded common-emitter stages without transformers. If a 5-stage amplifier is postulated, then the gain per stage (exclusive of the input stage) must be approximately 10. This is easily realized from cascaded common-emitter stages, and will still allow sufficient self-feedback for gain stabilization.

AMPLIFIER STABILIZATION

The first step in the stabilization process is determining the available fre-



$$Z_{io}(j\omega) = R_{io} \cdot \frac{1+j\frac{1}{\eta}\frac{\omega}{\omega_{cb}}}{1+j\frac{\omega}{\omega_{cb}}}$$

Fig. 9. Magnitude and phase characteristics of the short-circuit input impedance

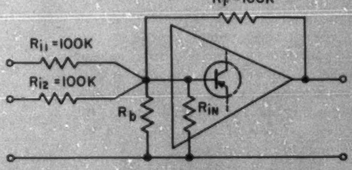


Fig. 11 (right). Design calculations for a 2-input amplifier.

$$R_b' = \frac{R_b R_{IN}}{R_b + R_{IN}} \qquad R_i' = \frac{\left(\frac{R_{i1} R_{i2}}{R_{i1}+R_{i2}}\right) R_b'}{R_b' + \left(\frac{R_{i1} R_{i2}}{R_{i1}+R_{i2}}\right)} \qquad \beta = \frac{R_i'}{R_i' + R_F}$$

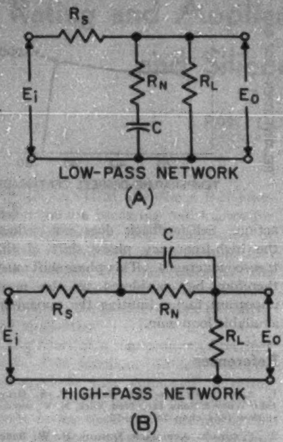
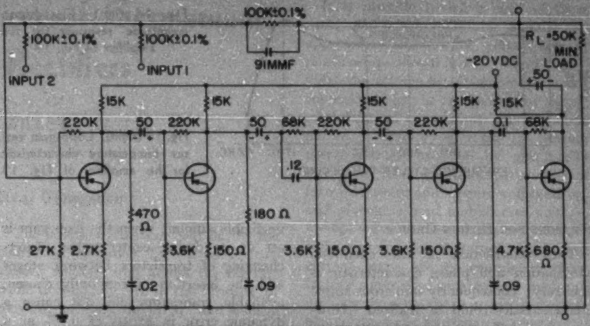


Fig. 12. Networks used for amplifier stabilization

A—Low-pass network $R_s=(\eta-1)R_n$
B—High-pass network

R_s = source resistance
R_n = network resistance
R_L = load resistance
η = frequency or ultimate attenuation ratio



TRANSISTORS 2N34, 2N109 (OR EQUIVALENT) K = X1000
ALL CAPACITORS IN MICROFARADS UNLESS OTHERWISE SPECIFIED
ALL RESISTORS 1/2 WATT ±10% UNLESS OTHERWISE SPECIFIED
MAXIMUM OUTPUT VOLTAGE INTO RATED MINIMUM LOAD IMPEDANCE 4.0V RMS

Fig. 13. Schematic diagram of an experimental 400-cycle computing amplifier

quency range in which the attenuation and phase may be controlled by interstage compensation networks. Except for impedance levels, the low-frequency stabilization problem is basically the same as in vacuum-tube amplifiers, and so will not be discussed.[10] However, as was shown, the phase shift of the transconductance will be the limiting factor at high frequencies, since it limits the frequency band in which effective compensation may be had to reduce the loop-gain below unity before the phase shift reaches −180 deg. A good engineering approximation is obtained if it is assumed that all transistors have the same g_{eb} gain and phase characteristics.

In the interests of simplicity and subminiaturization, the simple high-pass and low-pass networks of Fig. 12 are used for low- and high-frequency stabilization. Specification of the maximum network phase shift determines the ultimate attenuation ratio η, and therefore the ratio of network resistance to circuit resistance. The value of the capacitance fixes the network characteristics in a particular range of the frequency spectrum. Therefore, with the limiting attenuation and phase characteristics of the amplifier known, a stable loop-gain characteristic may be constructed. The results will dictate how many networks should be used, what their attenuation ratios should be, and at which frequencies they should be effective. The only additional information required to complete the design is the impedance level of the circuit in which the networks are to be placed. This procedure was followed in the present design and resulted in an unconditionally stable amplifier for the specified loop-gain of 1,000. Relatively few changes of component values were necessary to approximate closely the calculated $A\beta(\omega)$ characteristic. The high degree of prediction is attributed to the ability to determine the circuit impedance levels and gain in the presence of self-feedback.

The schematic diagram of an experimental 5-stage amplifier, based on the principles discussed thus far, is shown in Fig. 13.

Table I. Summary of Design Equations

Feedback	Input Impedance	Output Impedance	Current Gain	Voltage Gain
Voltage	$Z_{ivfb} \approx Z_{io} \parallel \dfrac{Z_f}{1+K_{vnfb}}$ $Z_f \gg Z_1$ $\alpha_{11} \gg 1$	$Z_{ovfb} \approx Z_{o\infty} \parallel \dfrac{Z_f}{1+\dfrac{\alpha_{11}Z_2}{Z_{in}}}$ $Z_f \gg Z_L$	$K_{ivfb} \approx \dfrac{K_{infb}}{1+\dfrac{Z_1}{Z_f}(\alpha_{11}+1)}$	$K_{vvfb} \approx \dfrac{K_{vnfb}}{1+Z^0\left[\dfrac{Z_1}{Z_2Z_f}+\dfrac{1}{Z_{ivfb}}\right]}$
Current	$Z_{icfb} \approx Z_{in}+(1+K_{infb})Z_e$ $Z_{o\infty} \gg Z_e$	$Z_{ocfb} \approx Z_{o\infty}\left[1+\dfrac{\alpha_{11}Z_e}{Z_e+Z_g+Z_{io}}\right]$ $Z_{o\infty} \gg Z_e$	$K_{icfb} \approx K_{infb}$ $Z_{o\infty} \gg Z_e$	$K_{vcfb} \approx \dfrac{K_{vnfb}}{1+\dfrac{Z_e}{Z_L}K_{vnfb}}$ $Z_{o\infty} \gg Z_e \quad K_{vnfb}Z_{io} \gg Z_L$
Total	$Z_{itfb} \approx Z_{ivfb}+(1+K_{ivfb})Z_e$	$Z_{otfb} \approx Z_{ovfb}\left[1+\dfrac{\alpha_{11}Z_e}{Z_g+Z_{ivfb}}\right]$	$K_{itfb} \approx K_{ivfb}$	$K_{vtfb} \approx \dfrac{K_{vvfb}}{1+\dfrac{Z_e}{Z_L}K_{vvfb}}$

Z_{io} = short-circuit input impedance with no feedback
$Z_{o\infty}$ = open-circuit output impedance with no feedback
K_{infb} = current gain with no feedback
K_{vnfb} = voltage gain with no feedback

α_{11} = short-circuit current gain
Z_e = series feedback impedance
Z_f = parallel feedback impedance
Z_L = load impedance

$$Z_1=\frac{Z_{o\infty}Z_L}{Z_{o\infty}+Z_L}$$

$$Z_2=\frac{Z_{in}Z_g}{Z_{in}+Z_g}$$

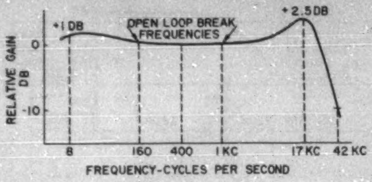
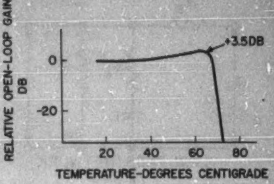


Fig. 14 (left). Closed-loop magnitude response for the amplifier of Fig. 13

Fig. 15 (right). Loop-gain versus temperature characteristic for the amplifier of Fig. 13



Reasons for Design Choice

The construction of the open-loop attenuation and phase characteristic is arbitrary; consequently numerous stable characteristics could be realized. However, it might be well to point out some of the reasons for the choice that was made. In an a-c computer utilizing many amplifiers in cascade it is important to keep the closed-loop phase shift small in the useful frequency band. This may be achieved by making the open-loop phase shift small in this range. Since the low- and high-pass stabilization networks introduce lagging and leading phase shifts respectively, their effects tend to cancel in the region about the carrier frequency. Ideally, if the network break frequencies had a geometric mean frequency exactly equal to the carrier frequency, exact cancellation would result. In the present design, this procedure was followed as far as possible. At frequencies considerably higher than the carrier frequency, transistor phase shift dominates, and thus carrier-frequency symmetry is lost. For the transistors used, the high-frequency phase shift was so severe that additional low-pass networks were no longer effective for gain-phase compensation. This problem is solved if a leading phase can be introduced at high frequencies which will partially offset the rapidly changing transistor phase shift. A cascade lead network is undesirable because of the accompanying low-frequency attenuation. Another means is making the β network exhibit phase lead properties by placing a capacitance in shunt with the feedback resistance. The minimum frequency at which the maximum phase lead may occur is dictated by how much residual phase shift is allowed at the carrier frequency, since this amount of phase shift will appear in the closed loop.

Test Results

Measurements on the circuit of Fig. 13 show that a stable loop gain of 60 db is obtained and that the static amplitude and phase errors may be reduced to a negligible amount when the loop gain is set at its design center value. Interchanging of transistors between stages and the insertion of randomly chosen, available transistors did not cause a dynamic error in excess of 0.1% at a signal input of 3.0 volts rms.

Fig. 14 is a plot of the closed-loop magnitude response of the circuit of Fig. 13. The magnitude of the resonant peaks gives an indication of the degree of stability. Fig. 15 shows the open-loop gain temperature characteristics. The rapid decrease above 65 degrees centigrade is attributed to the inability of self-feedback, using only pure resistance elements, to control I_{co} at these elevated temperatures. It is felt that this temperature characteristic could be improved considerably by using temperature-sensitive self-feedback elements such as thermistors and thyrite resistors.

Although exhaustive tests of the amplifier of Fig. 13 were not made, the limited results are encouraging, for this study indicates that high-accuracy analogue-computing amplifiers may be built with contemporary commercial transistors.

Conclusions

It was demonstrated that a high-accuracy amplifier may be realized by having a relatively low loop gain, provided the variation of the loop gain can be kept small. Since a high loop gain is not easily realizable with the transistors available for this investigation, it was imperative that the amplifier gain variations be minimized. This was accomplished by local current and voltage feedback or self-feedback. For the transistors used it was shown that self-feedback is a satisfactory means of reducing gain and terminal impedance variations. As a consequence of self-feedback, the gain and terminal impedances may be rather accurately computed with the use of manufacturers' data and the equations of Table I. Further, for a-c amplifiers, the undesirable effects of temperature are partially overcome by the self-feedback action. Self-feedback does not reduce the high-frequency phase shift of the transconductance. This phase shift must therefore be considered as the most important factor limiting the maximum available loop gain.

References

1. Applied Electronics (book), T. S. Gray. John Wiley & Sons, Inc., New York, N. Y., second edition, 1954, chap. IX, art. 12.

2. Feedback Amplifier Design, H. W. Bode. *Bell System Monograph no. B-1239*, Bell Telephone Laboratories, New York, N. Y. Vacuum-Tube Amplifiers, G. E. Valley, H. Wallman. *MIT Radiation Laboratory Series*, McGraw-Hill Book Company, Inc., New York, N. Y., vol. 18, 1948, pp. 383–83.

3. Some Circuit Properties and Applications of N-P-N Transistors, R. L. Wallace, W. J. Pietenpol. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 23, July 1951, pp. 753–67.

4. Wallace and Pietenpol. *op. cit.*

5. Application of Junction Transistors to Carrier Frequency Computing Amplifiers, W. A. Curtin. *S. M. Thesis*, Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass., Aug. 1955. pp. 22–33.

6. Curtin *op. cit.*, pp. 37–38.

7. Wallace and Pietenpol, *op. cit.*

8. Transient Analysis of Junction Transistor Amplifiers, W. F. Chow, J. J. Suran. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, Sept. 1953, pp. 1125–29.

9. A P-N-P Alloy Junction Transistor for Radio Frequency Amplification, C. W. Mueller, J. I. Pankove. *Ibid.*, vol. 42, Feb. 1954, pp. 386–91.

10. Valley and Wallman, *op. cit.*, pp. 370–76.

◆

Discussion

W. M. Grim, Jr. (General Electronics Laboratories, Inc., Cambridge, Mass.): The author has presented an excellent solution to the problem of designing a transistor amplifier with the gain stability which is required for computer use. The use of local feedback not only reduces the amount of over-all feedback necessary to maintain the required gain stability, but also reduces the effects of the transistor parameter variations on the feedback-loop stability.

The use of self-feedback techniques, together with circuits for stabilizing the operating conditions, may make the use of over-all feedback unnecessary, in applications where the gain stability requirements are less stringent.

Rating and Application of Germanium and Silicon Rectifiers

F. W. GUTZWILLER
ASSOCIATE MEMBER AIEE

GERMANIUM and silicon P-N junctions are becoming well known for their very nearly ideal combination of efficiency, flexibility, and reliability factors. Despite their high efficiency, the small internal losses of this type of rectifier cell are the chief criterion for rating and application.

The heart of a semiconductor rectifier is the P-N junction. In a philosophy analogous to that employed in rating practically every other type of electrical power device, the temperature rise of this P-N junction governs the successful application of the rectifier cell. Long life and high reliability can be expected if junction temperature is limited to a reasonable value under continuous duty conditions, and if an ample safety margin is maintained for unusual voltage and current transients that may be expected in service.

The temperature at which a junction may satisfactorily operate depends to a great extent on the design and fabrication of the rectifier cell. High temperatures promote chemical activity of contamination on the delicate and minute surfaces of the junction. This chemical activity deteriorates the voltage-blocking ability of the rectifier to the eventual point where breakdown or thermal runaway occurs.

Two obvious alternatives, or a combination of both, can protect the junction from this possibility. One is to limit junction temperature to such a low value that chemical activity is reduced to a negligible rate. The other, and more direct, approach eliminates the contaminating foreign agents from possible contact with the junction. This can be accomplished by extreme care in fabrication, chemical or electrolytic etching of the junction, rigorous cleaning processes (all to rid the device of initial contamination), and finally, a hermetic seal to keep subsequent environmental contamination away from the junction during the life span of the cell. Careful attention to these cleanliness factors permits trouble-free operation at junction temperatures in excess of 100 degrees centigrade for germanium cells and well over 200 degrees centigrade for silicon cells.

Heat Generation

The junction generates heat during the course of operation by virtue of two losses: voltage drop across the rectifier while it is conducting in the forward, or easy current-flow, direction; and leakage current through the cell when it is restraining the flow of load current in the reverse, or blocking direction. The forward voltage drop for a typical germanium rectifier cell for various load currents and junction temperatures is shown in Fig. 1. The leakage or reverse current for various reverse voltages and junction temperatures is shown in Fig. 2. The characteristic curves of silicon rectifier cells are similar in shape to those of germanium cells, differing primarily in over-all magnitude.

The heating of the junction for any particular forward current wave form, when averaged over a complete cycle, can by expressed as

$$P_f=\frac{\int_0^T v_f i_f dt}{\int_0^T dt} \tag{1}$$

where

P_f = average forward heating power
i_f = instantaneous load current, determined by input voltage wave form and circuit impedance
v_f = instantaneous forward voltage drop corresponding to i_f
T = period of one cycle

The average heating due to reverse voltage across the junction can be expressed as

$$P_r=\frac{\int_0^T i_r v_r dt}{\int_0^T dt} \tag{2}$$

where

P_r = average reverse heating power
v_r = instantaneous reverse voltage, determined by input voltage wave form
i_r = instantaneous leakage current corresponding to v_r and operating junction temperature

It is possible to solve both of these equations using suitable analytical approximations for the voltage-current relationships shown in Figs. 1 and 2, provided the voltage supply and load current wave shapes can be represented by simple mathematical expressions. As an alternative, satisfactory solutions for heating power can be attained by graphically integrating the instantaneous product of forward current and forward voltage drop, and reverse current and reverse voltage, for the wave shapes of the application. For even extremely distorted voltage and current wave shapes, completely satisfactory approximations can be made by dividing the wave shape up into ten or less equivalent-area sections in bar-graph fashion.

Total average heating at the junction (P_t) is the total of forward and reverse heating:

$$P_t=P_f+P_r \tag{3}$$

Silicon rectifiers, which have a substantially lower reverse current and a higher forward voltage drop than germanium rectifiers of equal power rating, have so little reverse heating compared to forward heating that P_r may be considered negligible in many cases.

Heat Dissipation

Fig. 3 is a cross section through a typical semiconductor power rectifier of the type shown in the 4-fin stack arrangement of Fig. 4. The P-N junction is soldered to a copper stud, which is in turn mounted on a radiator or cooling fin. This mechanism affords a highly conductive path to the cooling fin for the heat generated at the junction. The fin liberates this heat to the surrounding medium by radiation, convection, and conduction.

The flow of heat from the junction to a reference point on the fin, under stable conditions, can be expressed as

$$P_t=K_j(T_j-T_f) \tag{4}$$

where

K_j = thermal conductance between junction and fin
T_j = temperature of junction
T_f = temperature of fin

Likewise, the flow of heat from cooling fin to the surrounding ambient can be expressed as

$$P_t=K_f(T_f-T_a) \tag{5}$$

where

K_f = thermal coefficient of heat transfer between fin and ambient
T_a = temperature of ambient

Paper 56-934, recommended by the AIEE Metallic Rectifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted June 1, 1956; made available for printing July 20, 1956.

F. W. Gutzwiller is with the General Electric Company, Clyde, N. Y.

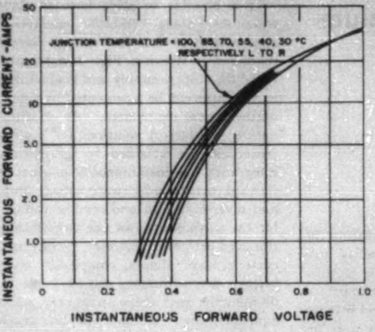


Fig. 1 (left). Forward characteristics of a typical semiconductor rectifier cell

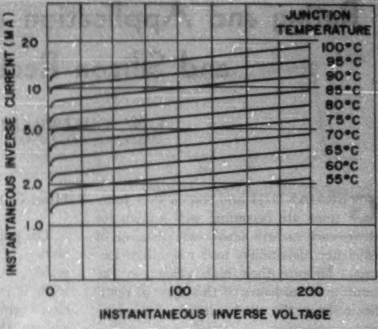


Fig. 2 (right). Reverse characteristic of a typical semiconductor rectifier cell

Combining these two equations to eliminate T_f, and introducing equation 3, gives

$$P_t=K_{eq}(T_j-T_a)=P_f+P_r \qquad (6)$$

where

$$K_{eq}=\frac{K_jK_f}{K_j+K_f}$$

This equation permits solution for the temperature rise between ambient and junction as a function of heating losses at the junction. Typical temperature rises for various cooling conditions (varying K_f), as a function of heating, are shown in Fig. 5. The preponderant effect of radiation when the rectifier is not forced air-cooled causes the curves for free convection air conditions to curve slightly from the linear form.

Rating the Cell

If this type of cooling curve, together with voltage-current characteristics and recommended junction operating temperatures, is available for a device, the equipment designer has the basic information for determining the continuous-duty operating temperatures of the rectifier junction under any conditions. Conversely, he has the data necessary to determine how many cells, and in what configuration he must arrange them, to achieve his load voltage and current requirements.

For commonly encountered types of voltage and current wave shapes, the device manufacturer usually publishes forward current ratings determined in the foregoing manner for various ambient temperatures and velocities of cooling air. A set of such curves for a cell used in a single phase sine wave voltage supply is shown in Fig. 6. Operation at any point of these curves develops the same junction temperature. Peak inverse voltage is maintained constant throughout.

In multiphase rectifier circuits, conduction of forward current through each leg is reduced to a shorter part of the whole cycle than in single-phase circuits, e.g., 120 degrees in an ideal 3-phase circuit. To attain the same average forward current, the cell must conduct instantaneously at a higher level. Since instantaneous heating increases at a rate proportionately greater than the increase of instantaneous forward current, the rectifier cell develops more average forward heat when operating in multiphase circuits. During the inverse part of the cycle, the cell must block voltage for two-thirds of the cycle in three phase, for example, instead of only one-half cycle as in single phase. For the same peak inverse voltage the 3-phase circuit will therefore develop additional reverse heating in the cell. Because of this more severe heating during conduction and blocking, semiconductor cells are usually derated for multiphase circuits, sometimes quite drastically.

Intermittent Duty

The foregoing approach determines the continuous-duty rating of a rectifier cell by calculation of the temperature rise between ambient and junction after power has been applied to the cell for an indefinitely long period of time. If, on the other hand, power is applied to the cell intermittently, transient heating occurs as heat is stored in, and then released from, the thermal capacity of junction, mounting, and cooling fin.

Essentially two temperature transients occur. If a cell is energized after having been allowed to stabilize at ambient temperature under no load and no voltage, the semiconductor junction very quickly ascends to its stable temperature rise above the cooling apparatus, in accordance with equation 4. This usually occurs in a matter of cycles, depending mainly on the thermal mass immediately surrounding the junction of the device, and the thermal conductance to the heat sink of the cooling apparatus.

The cooling system, which has a much larger mass than the junction, heats much more slowly, and may take as long as an hour to stabilize at its ultimate differential temperature above the ambient, in accordance with equation 5.

When a cell is de-energized, the junction temperature descends to the heat sink temperature within a few cycles. The heat sink, however, drifts only gradually down to ambient temperature as it releases its substantial amounts of stored heat to the cooling medium.

Depending on the application, both of these heating transients may be utilized to develop substantial improvements over continuous-duty ratings for intermittent types of operation, such as circuit-breaker closing and hoist drives. A circuit-breaker closing application demands rectified current for only the few cycles required to close the breaker. The cooling fin acts as a true heat sink in this case, and very large currents can be rectified by taking advantage of the short-term thermal capacity of the junction. A hoist drive, on the other hand, requires power a good deal longer than a few cycles, but it still can be considered an "on-off" type of load. In this case, the cooling fin has time to store substantially more heat while the load is imposed, and its temperature may rise considerably, but still a good deal less than if allowed to reach its stable temperature under continuous load. The magnitude of allowable loading for various duty cycles may be determined quite readily by analytical approaches.

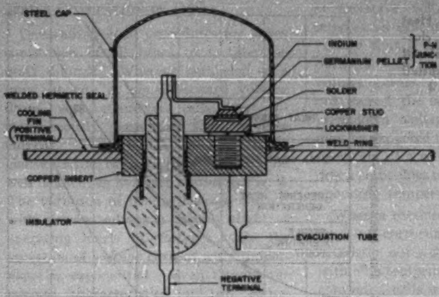


Fig. 3 (left). Cross section of a medium-power germanium rectifier cell

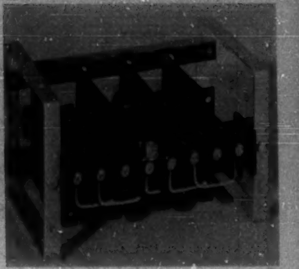


Fig. 4 (right). Typical rectifier stacking arrangement. Cooling fins are 3½ inches square

Thermal Stability

In well-fabricated germanium rectifier cells, the variation of reverse heating for junction temperatures within the normal operating range and for a fixed reverse voltage can be empirically represented by this equation

$$P_{rj} = P_{ra}\epsilon^{\gamma(T_j - T_a)} \tag{7}$$

where

P_{rj} = reverse heating at junction temperature T_j
P_{ra} = reverse heating with junction at ambient temperature T_a
γ = constant of temperature variation
ϵ = base of natural logarithm

This equation is plotted in Fig. 7. The heat generated by a specified forward current is relatively independent of junction temperature, dropping only slightly as the temperature is raised. This variation P_f is also plotted in Fig. 7, and is added to the reverse heating to yield total junction heating P_t. The equation of the total heating curve, considering forward heating constant with temperature, is

$$P_t = P_f + P_{rj} = P_f + P_{ra}\epsilon^{\gamma(T_j - T_a)} \tag{8}$$

The origin of the abscissa of Fig. 7 is the ambient temperature of the application. From this point, following the locus of equation 6, can be laid out the heat-dissipating characteristic of the rectifier.

When the rectifier cell is energized in the circuit, the junction will heat as defined by the heat generation curve, and will cool as defined by the heat dissipation curve. The junction temperature will rise to the point where heat generation exactly matches heat dissipation. This occurs where the two curves intersect. The area enclosed by the two curves is a rough measure of the inherent stability of the application.

If junction heating is increased by additional voltage or current, or by raising ambient temperature, or if cooling is reduced, the overlap of the two curves decreases, and, if taken to extremes, the two curves fail to intersect. Under these conditions, the rectifier cannot stabilize thermally. The temperature of the cell increases to the melting point of the junction or to the point where thermal stress causes cell failure. This action is called "thermal run-away."

An analytical solution for the critical operating point at which thermal run-away becomes imminent can be derived for a rectifier that meets the conditions of equations 6 and 8.

$$P_{ra\,\max} = \frac{K_{\theta a}}{\gamma}\epsilon^{-\left(1+\frac{\gamma P_f}{K_{\theta a}}\right)} \tag{9}$$

This critical value of reverse heating is measured with the junction at ambient temperature, the most readily available reference point. In actual application, substantial safety factors are usually applied to this critical value to allow for overloading, surges, and transient conditions, and for possible variations in the reverse characteristic of the rectifier cell due to contamination. Because of their inherently higher reverse heating, thermal stability criteria are usually of more importance and interest for germanium rectifiers than for silicon devices.

Some Circuit Design Considerations

Fusing

Under load fault conditions, as well as during rectifier faults, many times normal current flows through the cells. The resulting current densities are much higher than normal, and heating of the junctions becomes severe. Due to the limited thermal capacity of a junction, temperature rises quickly. Under conditions where circuit impedance is extremely low, very high fault currents result, and junction temperature may rise to the breakdown point within a half-cycle. To safeguard rectifiers under these circumstances, additional circuit impedance in the form of high transformer reactance or lumped inductance or resistance can be introduced to the circuit to limit fault current to the rectifier surge rating recommended by the manufacturer, until circuit protection can function. As a satisfactory alternative where voltage regulation requirements do not tolerate introducing extra impedance, current-limiting fuses can be used. These devices can open the faulted circuit in less than one-quarter cycle, before current can rise to its peak value and destructive proportions.

Surge Voltage Protection

When a semiconductor rectifier feeds an inductive load from a highly reactive a-c line, special protection against induced voltage surges may be necessary. Under these conditions, the rectifier cell conducts an essentially square wave of current. When the cell suddenly commutates its load to the next leg, a charge of current carriers are left stranded in the bulk semiconductor material. These carriers, known as "holes," normally disappear by recombination with electrons or by diffusion out of the semiconductor material. However, when reverse voltage is applied to the cell immediately after heavy forward current conduction, these carriers do not have time to recombine and diffuse naturally, but are instead swept across the junction into the P region, resulting in a current limited only by circuit impedance. Until the carriers have been completely swept out of the bulk semiconductor, the resistance of the cell is very low, acting almost like a short circuit.

Within a few micro-seconds after reverse voltage has been applied, the carriers have all been swept out, and the cell recovers abruptly. Its resistance immediately changes from a fraction of

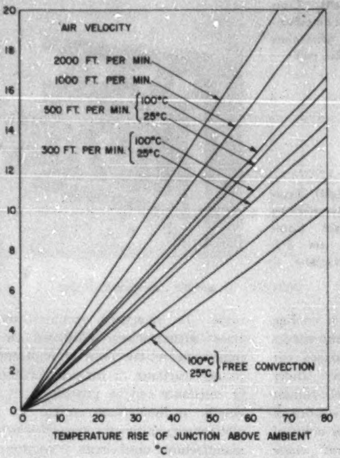
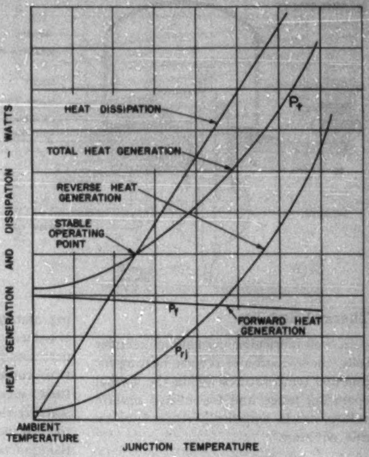


Fig. 5 (left). Heat dissipation characteristics of cell type illustrated in Fig. 4

Fig. 7 (right). Thermal characteristics of a rectifier cell under constant load and voltage conditions



an ohm to thousands of ohms, causing a substantial voltage surge to be induced in the line reactance. Particularly since this is a cyclical occurrence, it is important to protect the cells from these voltages if they are excessive. A few microfarads of capacitance installed across the a-c input terminals of the rectifier are usually sufficient to dampen the most vicious voltage spikes to values within rectifier capabilities, even in very high current installations.

A similar type of voltage surge may occur if fuses are used in individual legs of a rectifier circuit. In this instance, the voltage spike is not cyclical, but is generated in the line reactance when a fuse suddenly interrupts a high fault current. Here again, a small amount of capacitance across the a-c rectifier terminals will safeguard the cells from excessive voltage. Likewise, capacitors may be necessary to afford protection against line reactance voltage surges when load switching causes abrupt changes of current in the line reactance.

Paralleling Cells

The forward voltage-current relationship depicted in Fig. 1 illustrates the wide variation in forward current with relatively small changes in forward voltage drop. When cells are paralleled for higher current capacity, the voltage across the cells is necessarily equal, and it follows that the characteristics of the cells must be extremely well matched if they are to share current reasonably well. For this reason, the cell manufacturer usually prefers to perform this matching operation in his factory. To maintain equal current-sharing throughout the operating range of current and temperature, and under fault conditions, connections between cells must have equal resistance and be symmetrical, and connector length, cross section of wire, and type of connecting lugs should be maintained as nearly equal for each parallel branch as is practicable. Fig. 8 illustrates an effective manner of connecting cells or stacks in parallel.

This circuit can be used for adequately paralleling up to six or more matched cells; however, it is not the ultimate for current sharing. Further improvement can be made by bringing the external leads to the center cell on each of the respective stacks.

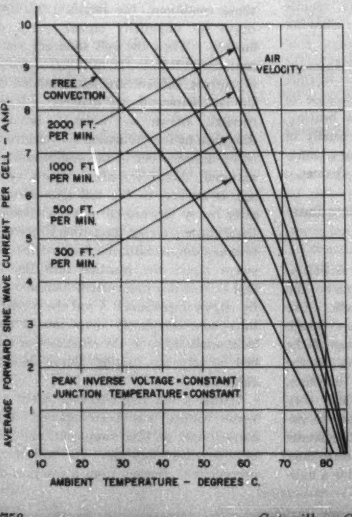


Fig. 6. Constant junction temperature ratings of a germanium rectifier cell

SERIES CONNECTION OF CELLS

Generalization of procedures for series connection of all types of rectifiers is quite difficult. Whereas paralleled cells must operate at equal forward voltage drops, series-connected cells must instantaneously pass the same reverse current. The reverse voltage-current characteristic of a rectifier cell can be well defined, as has been shown in Fig. 2. The variation of reverse leakage current with reverse voltage is usually slight at operating levels of voltage, while the variation of leakage current with temperature is exponential. Even though the reverse characteristics of two series-connected cells may be matched perfectly at the same junction temperature, a difference in junction temperature of only a few degrees will cause a grossly unequal distribution of voltage between the cells. Such a temperature variation must be expected in practical installations.

In some devices, depending on their application, the reverse heating of the cell is a large part of the cell dissipation. Under these conditions, substantial equalization of voltage between series-connected cells will be achieved, due to thermal self-correction. The cell that tends to assume a larger share of the voltage will heat up accordingly. Its reverse leakage current will tend to increase but, being limited by the other cells in series, the cell must decrease its share of the reverse voltage to normal values.

Other rectifier devices, particularly of the silicon type, exhibit a sharp nondestructive breakdown in reverse current at excessive voltage. This mechanism imposes an inherent limit on the reverse voltage that a cell can assume from its partners. As long as the total reverse voltage across all cells in series does not exceed the sum of their respective breakdown voltages, reliable operation may be expected from the cells, provided operation of a cell at, or near, its breakdown point does not exceed its maximum operating temperature limits.

When a rectifier cell exhibits neither a large self-correcting reverse heating characteristic, nor a sharp reverse breakdown within its temperature capabilities, series-connected cells will show substantial inequality in voltage distribution. To determine the effect of higher-than-normal reverse voltage on cell reliability, comprehensive life testing was conducted on medium power germanium rectifiers. For this particular type of device, no adverse effects were experienced on series-connected cells across which higher-than-rated peak inverse voltage was developed. The average peak inverse voltage across the cells in each leg was maintained at the normal cell rating. On the basis of these data, voltage-equalizing resistors, which had previously been used, were eliminated from circuits using medium power rectifier cells connected in series.

In a test to develop further information on this type of operation, medium power germanium rectifier cells were installed in three legs of the high-voltage power supply of a 50-kw short-wave transmitter. The circuit supplies 11,700 volts direct current and about 12.5 amperes under peak modulation. Forty-two germanium rectifier cells are connected in series in each of the three legs of the circuit with no attempt to ensure inverse voltage-sharing except matching of the reverse characteristics of the cells. The average peak inverse voltage of 293 volts per cell exceeds the nominal rating of the cells by approximately 45 per cent. In the 3 months to date that this test has been operating, the circuit has functioned without any incident in spite of the large overvoltage on the cells. It is felt that germanium and silicon will bring new standards of reliability and long life to this type of high-voltage d-c power supply.

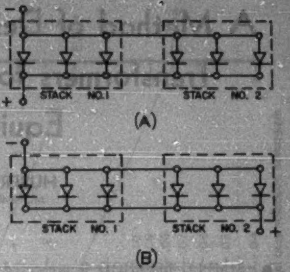


Fig. 8. Diagrams illustrating proper method of interconnecting rectifier cells and stacks

A—Unsymmetrical parallel connections: not recommended

B—Symmetrical parallel connections: recommended

A Method of Designing Small Power Transformers for Communication Equipment

HUBERT S. SEAR
NONMEMBER AIEE

A SIMPLE METHOD is presented for the design of a power transformer which does not require any trial-and error-calculations. Optimum wire gauges for each winding to minimize losses, and optimum cores made of commercially available laminations, can be easily obtained. The method is suitable for power transformers operating at any frequency where stray capacitative and inductive effects are negligible, having one or more secondary windings, and rated up to 500 volt-amperes.

Briefly, the window area, taken by each winding, is determined in terms of a factor K and one variable, N_p, the number of primary turns. The core cross-sectional area is also determined in terms of the factor k_p and of N_p. The product of these two factors determines the product of core cross-sectional area and window area. The derivation of this is shown in the appendix. Optimum core dimensions and lamination sizes are then selected from a nomogram on the basis of K and k_p.

Wire gauges can be selected easily which allow allocation of equal window area to primary and secondaries, thus minimizing copper losses. An actual example will serve to illustrate the design procedure, after which the condensed instructions in the summary should suffice for practical use.

EXAMPLE

Design a transformer to the following ratings.

Core Material: Allegheny Ludlum transformer *A*

Primary: 115 volts, 60 cycles

Secondary:

No.	Volts	Amperes	Volt-Amperes
1	640	0.150	96
2	5	3.0	15
3	6.3	3.0	18.9
			129.9

Insulation: 5-volt winding insulated for 1,300 volts to core

Current density: 1,000 circular mils per ampere

PROCEDURE

1. Estimated losses, 14 volt-amperes. Total volt-ampere rating 129.9+14=143.9 or 144.

2. Evaluate k_p for this transformer.

$$k_p=\frac{E_p\times 10^{-8}}{4.44\,fB_{max}k_{sf}}$$

where

E_p = primary voltage, 115 volts
f = frequency, 60 cycles
B_{max} = maximum flux induced in core. For Allegheny Ludlum transformer-grade silicon-steel laminations 70,000 lines per square inch is feasible
K_{sf} = stacking factor of laminations, 0.90
$k_p = 686$

3. Evaluate K for this transformer.

$$K=\frac{1}{t_p}+\frac{k_{s1}}{t_{s1}}+\frac{k_{s2}}{t_{s2}}+\frac{k_{s3}}{t_{s3}}$$

K_s = factor due to each secondary winding $=\frac{\text{secondary voltage}}{\text{primary voltage}}\times$ regulation (estd.)

$$k_{s1}=\frac{640}{115}\times 1.05=5.85$$

$$k_{s2}=\frac{5}{115}\times 1.10=0.0478$$

$$k_{s3}=\frac{6.3}{115}\times 1.10=0.0603$$

T_p = turns per square inch (including layer insulation) in primary

T_s = turns per square inch (including layer insulation) in each secondary

Evaluation of these factors requires selection of wire gauge to be used. This is done as in conventional designs, on the basis of current density desired, (circular mils per ampere).

$$\text{Primary current }\frac{144\text{ volt-amperes}}{115\text{ primary voltage}}=1.25\text{ amperes}$$

	Required Circular Mils	Wire Gauge to Be Used, No.
Primary	1,250	19
Secondary:		
1	150	28
2	3,000	15
3	3,000	15

Table I shows turns per square inch including paper layer insulation sufficient to support the wire. This wire table also shows the reciprocal of turns per square inch for easy substitution in the equation to evaluate K.

Primary *1:* $\frac{1}{T_p}$ $19.57\times 10^{-4}\times 1 = 19.57\times 10^{-4}$

Secondary *1:* $\frac{k_{s1}}{T_{s1}}$ $2.41\times 10^{-4}\times 5.85 = 14.10\times 10^{-4}$

Secondary *2:* $\frac{k_{s2}}{T_{s2}}$ $46.08\times 10^{-4}\times 0.0478 = 2.20\times 10^{-4}$

Secondary *3:* $\frac{k_{s3}}{T_{s3}}$ $46.08\times 10^{-4}\times 0.0603 = 2.78\times 10^{-4}$

$K = 38.65\times 10^{-4}$

4. The product of $K\,k_p$ would now be obtained and used on Fig. 1. However, since so many transformers are designed for 115-volt 60-cycle operation and use 4 per-cent silicon-steel cores, the nomogram in Fig. 2 has been developed requiring only evaluation of K. This limitation also allows inclusion of a scale on the nomogram giving the number of turns in the primary winding (N_p). The nomograms relate the core dimensions to k or $K\,k_p$. A_t is the total window area available for windings leaving insulation space indicated in Fig. 3. A_c is the cross-sectional area of the core. N_p is the number of primary turns. Optimal transformers have a square-core cross section and have the window completely filled.

Some commonly used lamination types are indicated on the nomograms by lines connecting their available window area with the cross-sectional area resulting in a square center leg. Therefore, if a square cross section is desired for a particular $K\,k_p$ value (or K value), a ruler is rotated about this value on the K scale until it intersects the A_c value for one of the lamination types. The intersection of the ruler with the A_t scale will give the window area required for the transformer. If this window area exceeds that available in the lamination type chosen on the A_c scale, this lamination type will not be suitable when a square center leg is used. The ruler is then rotated about the same point on the K or $K\,k_p$ scale until either

---

Paper 56-969, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted May 9, 1956; made available for printing August 31, 1956.

HUBERT S. SEAR, M.D., resides in Waltham, Mass.

Table I. Wire Table

Brown and Sharpe Gauge	Circular Mils	Single Enamel Wire (Corrected for Space Factor Shown) Turns Per Inch	Single Enamel Wire (Corrected for Space Factor Shown) Space Factor	Thickness of Paper Layer Insulation	Turns Per Square Inch	Reciprocal of Turns Per Square Inch
44	4.00	.369	.85	.0005	131,786	0.7588 $\times 10^{-5}$
43	4.84	.340	"	"	113,333	0.8824 $\times 10^{-5}$
42	6.25	.304	"	"	89,412	1.118 $\times 10^{-5}$
41	7.84	.265	"	.0007	67,949	1.472 $\times 10^{-5}$
40	9.61	.239	.86	"	55,581	1.799 $\times 10^{-5}$
39	12.25	.215	.86	"	45,745	2.186 $\times 10^{-5}$
38	16.00	.193	.87	.001	34,464	2.902 $\times 10^{-5}$
37	20.30	.170	"	"	27,869	3.588 $\times 10^{-5}$
36	25.00	.155	"	"	23,134	4.323 $\times 10^{-5}$
35	31.40	.140	.88	"	18,919	5.286 $\times 10^{-5}$
34	39.70	.124	"	"	15,122	6.613 $\times 10^{-5}$
33	50.40	.110	"	.0013	11,828	8.455 $\times 10^{-5}$
32	64.00	98	"	.0013	9,515	10.51 $\times 10^{-5}$
31	79.20	88	"	.0015	7,652	13.07 $\times 10^{-5}$
30	100	80	.89	"	6,349	15.75 $\times 10^{-5}$
29	128	71	"	"	5,071	19.72 $\times 10^{-5}$
28	159	64	"	"	4,156	24.06 $\times 10^{-5}$
27	202	57	"	.0022	3,220	31.05 $\times 10^{-5}$
26	253	52	"	"	2,680	37.31 $\times 10^{-5}$
25	320	47	.90	"	2,186	45.74 $\times 10^{-5}$
24	404	42	"	"	1,765	56.67 $\times 10^{-5}$
23	511	37	"	.005	1,267	78.93 $\times 10^{-5}$
22	645	33	"	"	1,028	97.28 $\times 10^{-5}$
21	812	30	"	"	852	117.4 $\times 10^{-5}$
20	1,020	26	"	.005	667	0.001499
19	1,300	23	"	.007	511	0.001957
18	1,600	21	"	"	429	0.002331
17	2,030	19	"	"	352	0.002841
16	2,600	17	"	.010	270	0.003704
15	3,250	15	"	.010	217	0.004608
14	4,100	13	"	"	171	0.005848
13	5,180	12	"	"	147	0.00680
12	6,530	10	"	"	110	0.00908
11	8,235	9	"	"	87	0.01150
10	10,380	8	"	"	70	0.01429

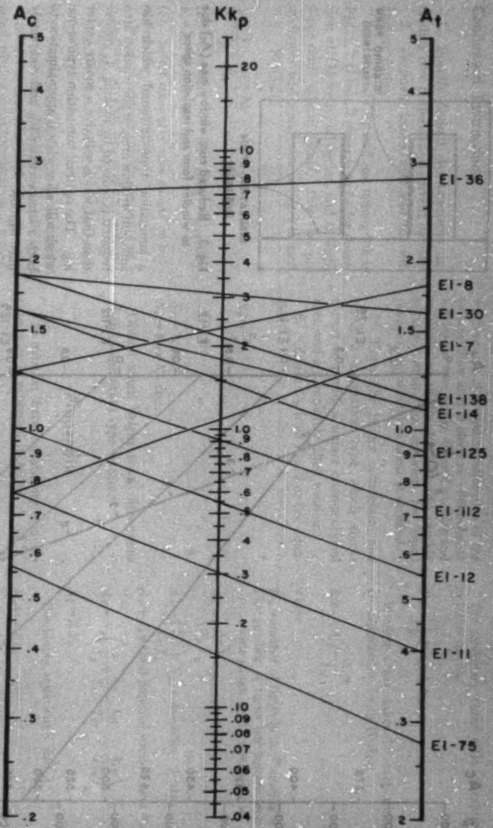


Fig. 1. Universal core selection nomogram

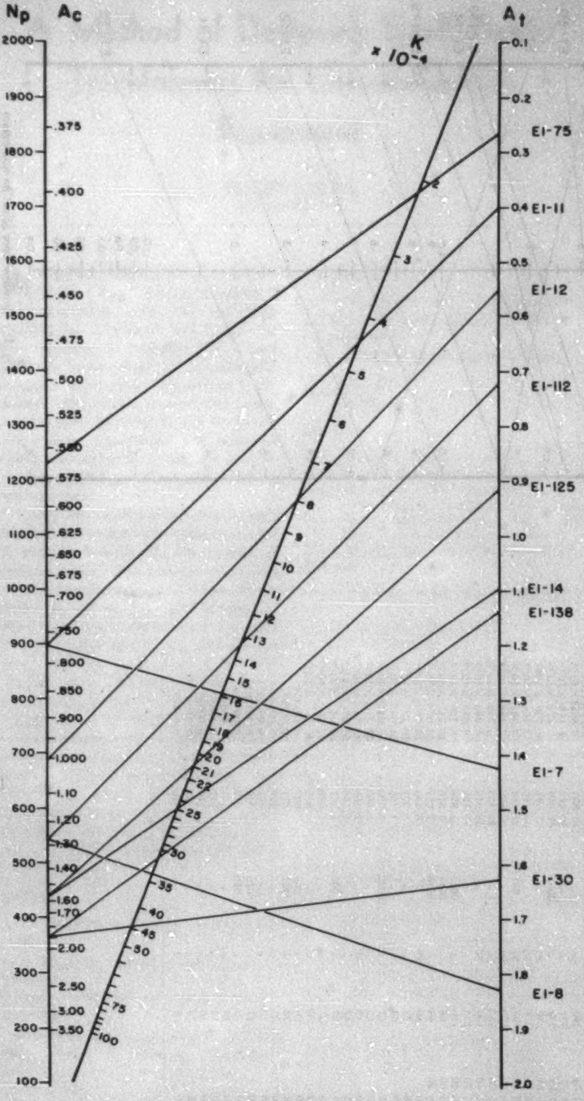


Fig. 2. Special core selection nomogram

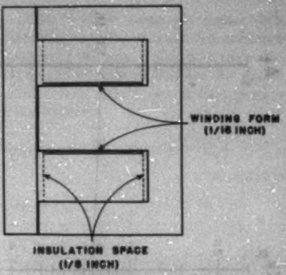


Fig. 3. Reduced core window area (A_t) due to winding form and insulation space

a lamination type is found which has sufficient window area, or the window area required is reduced by making A_c greater than that which will give a square center leg. The optimal lamination type is one which will allow closest approximation to both a square center leg and completely filled window.

By entering the nomogram in Fig. 2 at $K=39\times10^{-4}$ it is found that the line for lamination-type *EI-30* lies very close. However, if a square cross section at 1.90 square inches is taken, the window will not be completely filled, while a completely filled window (with useable area of 1.62 inches) will not allow a square cross section for the core. The nomogram allows selection of the optimal combination. It should be observed that, although the core cross-section area can be varied above and below the value indicated by the intersection of the lamination line with the A_c scale by varying the stack height, the window area A_t available is not so flexible. The value, indicated on the nomogram, may be exceeded only slightly at the expense of insulation spaces, marked in Fig. 3.

Taking the window area, which would be filled as 1.6 square inches, a core cross-sectional area of 1.7 square inches about 410 primary turns is obtained. Window space has been left for the high-voltage insulation around secondary 2. Then secondary turns are obtained from Np and k_s ($N_s=k_s\times N_p$)

Secondary *1:*	$5.85\times410=2{,}400$
Secondary *2:*	$0.0478\times410=19.6$ or 20
Secondary *3:*	$0.0603\times410=24.7$ or 25

Stack height is obtained by dividing the area obtained, by the center leg width given in Allegheny Ludlum Steel Corporation catalog *EM3* for lamination type *EI-30*.

$$\text{Stack height}=\frac{1.70}{1.375}=1.24 \text{ inches}$$

Condensed Instructions

1. Specify desired characteristics, including core material and current density.

2. Estimate losses, and add to volt-ampere rating.

3. Evaluate k_p, and use nomogram in Fig. 1. If transformer is to be for operation at 115 volts, 60 cycles, and utilizing 4-per-cent silicon steel, this is not necessary and nomogram in Fig. 2 is used.

4. Evaluate

$$K=\frac{1}{T_p}+\frac{k_{s1}}{T_{s1}}+\frac{k_{s2}}{T_{s2}}\cdots\cdots\frac{k_{sn}}{T_{sn}}$$

(*a*). Evaluate k_s for each secondary. $k_s=E_s/E_p\times$regulation.

(*b*). Select secondary wire gauges on basis of desired current density.
Calculate primary current (=total volt-amperes divided by primary voltage), and select required wire gauge.
Obtain reciprocal of turns per square inch from wire table.

(*c*). Substitute results of steps (a) and (b) to obtain value of K. It should be noted that losses are minimized by allocating equal window areas to primary and secondary windings. Since each fraction in the equation for K represents the relative window area of a winding, allocation of the total window area is obvious. Selection of a different wire gauge for one winding and recalculation of the fraction will quickly indicate when the criterion for equal primary and secondary areas is satisfied.

5. Enter the proper nomogram, and select a lamination type which comes closest to giving a square cross-sectional area and a completely filled window.

6. Obtain N_p from Fig. 2 or the equation

$$N_p=\frac{k_p}{A_c}$$

Obtain

$$N_{s1}=k_{s1}\times N_p$$
$$N_{s2}=k_{s2}\times N_p \text{ etc.}$$

Obtain stack height from dimensions of the lamination type selected.

Appendix

The relation between primary winding quantities is expressed in terms of N_p.

$$N_p=\frac{E_p\times 10^{-8}}{4.44fB_{max}\,AcK_{sf}}=\frac{k_p}{A_c} \tag{1}$$

The secondary turns ratio is expressed in terms on N_p.

$$N_s=\frac{Es}{Ep}N_p r=k_s N_p \tag{2}$$

Window areas, taken by each winding, are expressed in terms of N_p.

$$\text{Area}=A_n=\frac{N_n}{T_n} \tag{3}$$

where

N_n = number of turns in any winding n
T_n = number of turns per square inch for wire gauge used in winding n

$$A_p=\left(\frac{N_p}{T_p}\right)=\text{primary area}$$

$$A_{s1}=\left(\frac{k_{s1}}{T_{s1}}\right)(N_p)=\text{secondary } 1 \text{ area}$$

$$A_p+A_s=\text{total required window area}=A_T$$

$$A_T=N_p\left(\frac{1}{T_p}+\frac{k_{s1}}{T_{s1}}+\frac{k_{s2}}{T_{s2}}\cdots\frac{k_{sn}}{T_{sn}}\right)$$

$$A_T=KN_p$$

Combining equations 1 and 3

$$A_cA_t=Kk_p \tag{4}$$

The nomograms allow easy solution of this equation.

Accurate Determination of the Capacitance of a Thin Rectangular Plate

DANIEL K. REITAN
MEMBER AIEE

THOMAS J. HIGGINS
MEMBER AIEE

THE PURPOSE of this paper is two-fold: to determine by the method of subareas the charge distribution and capacitance of a rectangular area; and to advance therefrom a universal curve which furnishes accurate values of the capacitance of a rectangular area of any specified dimensions. From well-known theory, rigorous determination of these electrical quantities hinges on determination of the corresponding electrostatic potential function V. However, determination of the function for a specified plane area is of such mathematical difficulty that solution to date has been achieved for only one plane area, the elliptical disk of which the circular disk is a special case. In consequence, explicit equations for the charge distribution and electrical capacitance are known only for these two plane areas. However, need of accurate knowledge of the capacitance of a rectangular plane area, or of a certain parameter[1] equal in magnitude to the capacitance, occurs in numerous physical problems. (Thus, the normal force R on a thin rectangular plate moving in the direction of its normal with a velocity V through a liquid of viscosity μ is $R=8\pi\mu CV$, where C is a parameter equal to the capacitance of a rectangular planar conductor of the same dimensions as the thin plate.) To cite an example of much current interest in electrical engineering, a knowledge of the capacitance of rectangular plates is required in the rational design of grounding systems for high-voltage power systems.[2–4] Accordingly, it is obviously most desirable to have in hand curves which readily afford values of the capacitance with accuracy sufficient for design purposes. Precisely such a set of curves is advanced in this paper, the essential content of which may be summarized as follows: The method of subareas, earlier formulated and used by the authors to obtain excellent approximate solutions of the difficult, long-standing problems of determination of the capacitance and charge distribution of a plane annulus[5] and a cube,[6,7] is recast to enable solution of the present problem. Therewith equations for the capacitance and charge distribution of an arbitrarily dimensional rectangular area are formulated; and the solutions of these equations, as effected through use of International Business Machines equipment, are then utilized to plot a universal curve of capacitance versus width/height from which can be read the capacitance of any specified rectangular area with sufficient accuracy for most design purposes. Corroborative of the correctness of the solution in general, this curve falls within the narrow upper and lower bounds to be calculated from the known expressions for the capacitances of the circumscribed and

---

Paper 56-980, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted June 6, 1955; made available for printing July 24, 1956.

DANIEL K. REITAN and THOMAS J. HIGGINS are with the University of Wisconsin, Madison, Wisc.

This paper is based on a thesis supervised by Prof. Higgins and submitted by Prof. Reitan to the faculty of the University of Wisconsin, June 1951, in partial fulfillment of the requirements for the degree of Master of Science in electrical engineering.

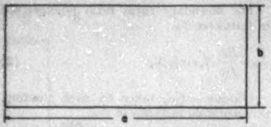
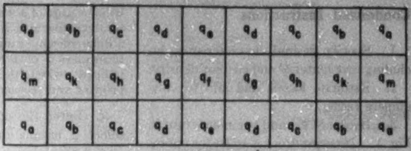


Fig. 1. Arbitrarily dimensioned rectangle with uniformly distributed charge density



q_a	q_b	q_c	q_d	q_e	q_d	q_c	q_b	q_a
q_m	q_k	q_h	q_g	q_f	q_g	q_h	q_k	q_m
q_a	q_b	q_c	q_d	q_e	q_d	q_c	q_b	q_a

Fig. 2. A 9 by 3 cm rectangle comprised of 27 unit subareas

inscribed ellipses. Finally, comparison of this universal curve with the corresponding curve plotted from Howe's[3] well-known and much used expression for the approximate capacitance of a rectangular area reveals that this expression yields values in considerable error over much of the range of w/h (width/height).

Basic Theory of Calculation

The basic theory of the determination of capacitance and charge density distribution by the method of subareas is detailed in references 5 and 6. Accordingly, it suffices here to recapitulate the essential steps of calculation. First, the particular area under consideration is divided into geometrically identical subareas, sufficiently small that the charge density is substantially constant over them and of such shape that the potential produced at any arbitrary point by a uniformly charged subarea can be calculated without too great difficulty. Second, at the same geometrically located point in each subarea the potential V_i due to its own charge and to the charges on the other subareas, assumed uniformly distributed over these subareas, is calculated. Taking into account any symmetry present in the subdivision, the procedure yields a set of n linear algebraic equations in terms of the n unknown charges on the subareas. Third, paralleling the known condition that the potential is the same on each subarea when the charge distribution is in equilibrium, the potential V_i on each subarea is set equal to V_0. Fourth, solution of the resulting set of n linear algebraic equations yields the charge on each subarea, as expressed in terms of the equipotential V_0. Fifth, summing these charges on each subarea yields the total charge Q, expressed in terms of V_0. Sixth, approximate numerical value of the capacitance is calculated from $C \cong Q/V_0$. Obviously, the approximate distribution of charge is afforded by the known values of the charges of the subareas.

The capacitance and charge density can be obtained to a high degree of accuracy by taking subareas sufficiently small in size. Of course, the labor involved in solving the set of n linear equations increases rapidly with n. However, as evidenced by the illustrative examples of this paper, surprisingly accurate values of capacitance and charge distribution can be obtained with use of small n, particularly if the area in question possesses symmetry about one or more axes.

Determination of Capacitance of a Thin Rectangular Plate

As a first approximation to calculation of the capacitance of a rectangle (thin rectangular plate) consider a charge distribution as in Fig. 1: the charge density is assumed constant over the surface of the rectangle. Then as detailed in Appendix I, the approximate value of capacitance to be computed with this distribution is $C = 0.597 \mu\mu f$ (micromicrofarad) for a rectangle of 3 by 1 cm (centimeters). The charge density, as used here, is the sum of the equal charge densities at corresponding points on the two opposite sides of the thin plate, hence twice the actual charge density on one side. In that the problem under discussion is one of absolute electrical measurement, it is convenient to work in unrationalized centimeter-gram-second units. Accordingly, the units of all quantities mentioned are in this system, except as specifically noted otherwise.

As a second approximation to calculation of the capacitance of a 3 by 1 cm rectangle, consider a charge distribution as in Fig. 2, comprising a 9 by 3 rectangle subdivided into 27 uniformly charged subareas. A well-known theorem in electrostatics states that the ratio of the capacitances of two geometrically similar areas is equal to the ratio of the corresponding linear dimensions. Thus, the capacitance of the 9 by 3 rectangle is three times that of the capacitance of a 3 by 1 rectangle. Appreciation and use of this theorem enables employment of the unit square as the basic subarea, with consequent simplification of numerical computation.

Consideration of symmetry limits the number of unknown q_i to ten. The potential V_a at the center of subarea A_a is comprised of a component of potential contributed by the charge q_a itself plus components contributed by each of the remaining subareas. Now, as detailed in Appendixes I and II of reference 6, the potential at the center of a unit subarea due to its own charge of uniform density q is 3.52549 q; and the potential at the center of a unit subarea due to the uniform

Vol. 7
Pt. 1
1956

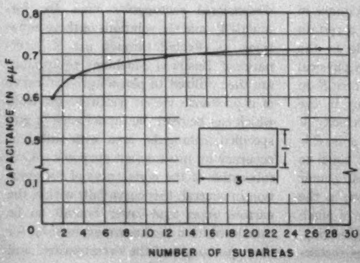
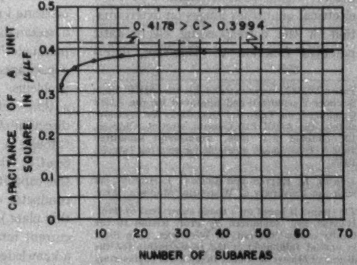


Fig. 3 (left). Capacitance of 3 by 1 cm rectangle as function of number of subareas

Fig. 4 (right). Capacitance of unit square as function of number of subareas



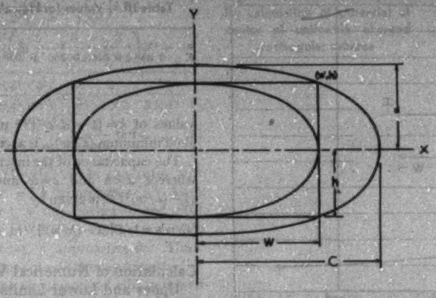


Fig. 5. Co-ordinates for calculation of upper and lower limits of capacitance of rectangle

density q' on any second subarea is well approximated by q'/d, where d is the distance between their centers.

Utilizing these relationships to determine the total potential at the center of a subarea of charge density q_a, and performing the necessary arithmetic, yields the first of the set of equations of Appendix II for the 9 by 3 rectangle. Proceeding similarly for subareas b through m yields the remaining equations. It follows from the geometric symmetry of the rectangle that the coefficients k_{ij} and k_{ji} $(i \neq j)$ either should be equal or one should be twice or four times the other. The corresponding coefficients of the set of equations satisfy these conditions. Next, imposing the condition that all V_i equal V_0, and solving the resulting set of equations, yields the numerical values tabulated in Appendix II. In return, these yield the associated approximation Q_t to the total charge Q on the plate as

$$Q_t = \sum_{i=a}^{m} q_i \text{ area}_i = 1.92797 V_0 \tag{1}$$

Accordingly, the corresponding approximation to the capacitance $C = Q/V_0$ of a 3 by 1 rectangle is

$$C \cong 1.92797 V_0 / 3 V_0 = 0.64265 \text{ esu} = 0.71405\ \mu\mu f$$

where esu is electrostatic units.

Fig. 3, which comprises a plot of the capacitance of a 3 by 1 rectangle versus the number of subareas used in obtaining a corresponding value, is evidence of the excellent accuracy gained by use of only 27 subareas and of the fact that only a small increase in accuracy is to be gained by use of more subareas.

Special Case of a Square

Similar calculation for the particularly interesting configuration of the square yields the convergence curve of Fig. 4. Extrapolation of the subarea values according to the method delineated by Allen and Dennis,[9] yields a limiting value $C = 0.4019\ \mu\mu f$. Corroboratively, this value is in essential agreement with the value of 0.40 $\mu\mu f$ given by Maxwell.[10,11] Also it falls within the limits of $0.4178 > C > 0.3994$ $\mu\mu f$, given by Polya and Szegö[12] through use of Schwarz's principle of symmetrization. It is of interest to note that the lower limit 0.3994 $\mu\mu f$ was rigorously derived by Lord Rayleigh; however, the upper limit is based on Polya and Szegö's conjecture that of all surfaces that can be enclosed by an outermost radius r, a circular plate of this same radius r has the maximum capacitance.

Limits on Possible Value of Capacitance of a Rectangle

As is evident from Fig. 4, application of the method of subareas to effect the approximate capacitance of the special case of the square yields fair accuracy on a basis of 16 increments and excellent accuracy on a basis of 36 increments. Correspondingly, it may be inferred that the solution of the 3 by 1 rectangle on a 12-increment basis is less than the exact value, whereas the solution on a 27-increment basis ought to be close to the exact value. Polya and Szegö's bounds ascertain the explicit accuracy of a subarea value for the square. On the other hand, inference as to the relative accuracy of the 27-increment subarea value for the 3 by 1 rectangle can only be conjectured in the absence of similar bounds for a rectangle. The obviously desirable effectance of these bounds will now be discussed.

Consider the rectangle and the corresponding inscribed and circumscribed ellipses of Fig. 5. By virtue of the well-known theorem in electrostatic theory which states that the capacitance of a conductor is greater than the capacitance of any part of it, it follows that the capacitance of the rectangle is greater than the capacitance of the inscribed ellipse and less than the capacitance of the circumscribing ellipse. Inasmuch as the dimensions of the rectangle fix the dimensions of the inscribed ellipse, the lower limit thus obtained is definitely fixed in value. On the other hand, ellipses of

Table I. Normalized Ratio *C/w as Function of w/h

w/h	$(C/w)_i$	$(C/w)_{em}$	Average	C/w, Subarea	C/w Rayleigh's, Lower Limit†
1	0.798	1.000	0.853	0.804	0.798
2	0.515	0.724	0.620		0.564
3				0.476	0.461
4	0.397	0.550	0.473		0.399
6	0.347	0.478	0.412		0.326
10	0.301	0.407	0.353		0.252
10^2			0.184		
10^3			0.133‡		
10^4			0.104		
10^5			0.085		
10^6			0.073		
10^{10}			0.045		

*Values of C are in microfarads; values of w in centimeters.
†Rayleigh's lower limit: Capacitance of a rectangle $> (2\pi^{-3/2}a)$, where a is the side of a square of equal area.
‡For $w/h > 10^3$, $C/w \cong \dfrac{1}{Ln \dfrac{4w}{h}}$.

Table II. Capacitance Versus Number of Subareas, $\mu\mu f$

Number of Subareas	C 3×1 Rectangle	C Square
1	0.597	0.315
3	0.647	
4		0.356
9		0.373
12	0.696	
16		0.382
27	0.714	
36		0.392
∞		0.402

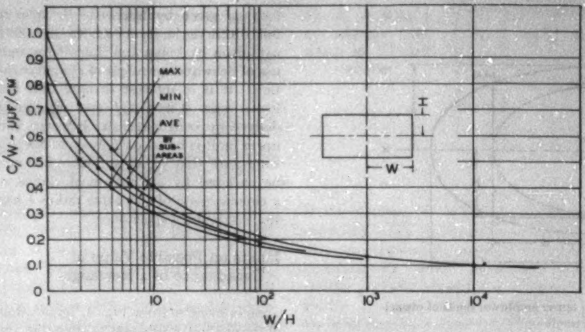


Fig. 6. Normalized plot of upper and lower limits of capacitance of rectangle

Table III. Values for Howe's Constant

α	1	2	3	5	10	20
K	0.336	0.340	0.360	0.388	0.448	0.582

various eccentricity can be circumscribed about the rectangle, whence determination of the minimum upper limit hinges on ascertaining the circumscribing ellipse having minimum capacitance. This latter can be obtained as follows.

The capacitance (see reference 10, page 237) of the circumscribed ellipse is $C=c/F(k)$ where c is the length of the major semiaxis, $k=[1-(d/c)^2]^{1/2}$ is the eccentricity, and $F(k)$ is the elliptical integral of the first kind with argument k. Inasmuch as the point (w,h) of Fig. 5 lies on the circumscribed ellipse whose equation is $(x^2/c^2)+(y^2/d^2)=1$, it follows that $(w^2/c^2)+(h^2/d^2)=1$. Multiplying each member by c and substituting $1-k^2=(d/c)^2$ yields $c^2=w^2+[h^2/(1-k^2)]$, whence

$$c=w[(1-k^2)+(h/w)^2]^{1/2}/(1-k^2)^{1/2} \qquad (2)$$

Substituting this value of c in $C=c/F(k)$ yields the capacitance of the circumscribing ellipse as

$$C=w[(1-k^2)+(h/w)^2]^{1/2}/(1-k^2)^{1/2}F(k) \qquad (3)$$

Dividing each side of equation 5 by w yields

$$(C/w)_c=[(1-k^2)+(h/w)^2]^{1/2}/(1-k^2)^{1/2}F(k) \qquad (4)$$

For a rectangle of specified ratio h/w, the ratio C/w for the circumscribing ellipse is a function of k alone. Hence, the value of k yielding the minimum value $(C/w)_{cm}$ of $(C/w)_c$ and thus of C when w is specified, can be found by the usual procedure for determining the minimum value of a function of one variable, i.e., differentiating $(C/w)_c$ with respect to k, setting the resulting expression equal to zero, and solving the equation thus formed for k_m. In that the equation which must be solved for k is a complicated transcendental equation involving elliptical functions, the simplest method of solving for the desired minimum value of $(C/w)_c$ is by substituting successive values of $k=[1-(d/c)^2]^{1/2}$ until the desired minimum $(C/w)_{cm}$ is ascertained.

The capacitance of the inscribed ellipse, whereof $d=h$ and $c=w$ and thus $k=[1-(h/w)^2]^{1/2}$, is given by

$$(C/w)_i=1/F\{[1-(h/w)^2]^{1/2}\} \qquad (5)$$

Calculation of Numerical Values of Upper and Lower Limits

When $w/h=1$, the rectangle is a square. The inscribed ellipse is a circle of radius $r=h=w$ and has a $(C/w)_i$ ratio of

$$(C/w)_i=1/F[(1-1)]=1/F(0)=1/1.571=0.637 \text{ esu}=0.708\ \mu\mu f \qquad (6)$$

This value, and other values of $(C/w)_i$ as determined by substitution of different w/h ratios in equation 5, are shown in Table I.

The values of $(C/w)_{cm}$, determined by successive substitution of various values of k in equation 4 for different w/h ratios, are also shown in Table I. It may be noted that the value for $w/h=1$, when the minimum circumscribing ellipse is a circle, is $(C/w)_{cm}=0.900$ esu $=1.00\ \mu\mu f$. Corroboratively, this value of 1.00 $\mu\mu f$ is in agreement with the value to be obtained from the well-known equation for the capacitance of a circular disc: $C=2r/\pi$. For in this case of $w=1$, the square is two units on a side, and the circumscribing circle is of radius $r=1.414$; whence $(C/w)=2.828/\pi=0.900$ esu $=1.00\ \mu\mu f$. The upper and lower limits for C/w when $w/h=2$, 4, 6, and 10, as shown in Table I, are obtained in similar fashion.

As the value of w/h becomes large (over 100), it is increasingly difficult to compute $(C/w)_{cm}$ by use of available numerical tables of elliptical functions. However, in this range the value of k is very close to unity and calculation is easily effected by use of the well-known series expansion[13] for $F(k)$

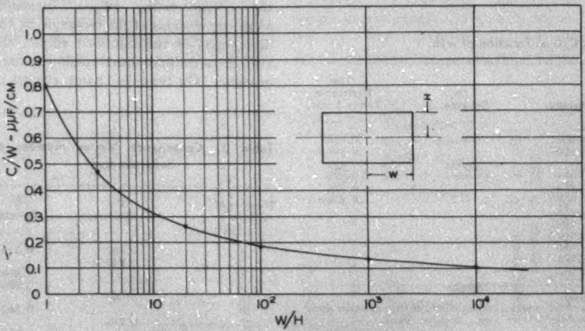


Fig. 7. Normalized curve of capacitance of rectangle

Table IV. Final Values for Universal Curve of Fig. 7

w/h	C/w	Basis for Computation
1	0.804	Subareas
3	0.476	Subareas
20	0.266	Howe's constant
10^2	0.184	$C/w \cong 1/(Ln\ 4w/h)$
10^3	0.133	$C/w \cong 1/(Ln\ 4w/h)$
10^4	0.104	$C/w \cong 1/(Ln\ 4w/h)$

$$F(k)=\ln\frac{4}{k'}+\frac{1^2}{2^2}\left(\ln\frac{4}{k'}-\frac{2}{1\times 2}\right)k'^2+\frac{1^2\times 3^2}{2^2\times 4^2}\left(\ln\frac{4}{k'}-\frac{2}{1\times 2}-\frac{2}{3\times 4}\right)k'^4+\frac{1^2\times 3^2\times 5^2}{2^2\times 4^2\times 6^2}\left(\ln\frac{4}{k'}-\frac{2}{1\times 2}-\frac{2}{3\times 4}-\frac{2}{5\times 6}\right)k'^6+\ldots \tag{7}$$

where $k'=(1-k^2)^{1/2}$. As the ratio w/h approaches ∞, k approaches 1. As k approaches 1, k' approaches 0, and $F(k)$ approaches $\ln (4/k')$. Further, as w/h approaches ∞, c approaches w. Thus, for large values of w/h the capacitance $C=c/F(k)$ of the circumscribed ellipse is well approximated from above by $C\cong w/\ln(4w/h)$. This is also the approximate value of the inscribed ellipse. Thus, for large values of w/h the upper and lower limits coalesce, both yielding an excellent approximation to the exact capacitance of the rectangle. Correspondingly, the values of C/w in Table I for ratios of $w/h>10^2$ are as computed from $C/w\cong 1/\ln(4w/h)$.

The curve of Fig. 6, plotted from the values of Table I, reveals at a glance the quality of the available limits on the possible value of the capacitance of a rectangle of specified ratio w/h. Also in Fig. 6 have been entered the nearly exact values of C/w for a square as cited, and for a 3 by 1 rectangle as calculated in Appendix II. On the basis of these points, it is to be concluded that the values of capacitance of rectangles of small w/h ratio are better approximated by the curves for the inscribed ellipse or for Rayleigh's lower limit than by the curve of averaged values of C/w for the circumscribed and inscribed ellipses.

Comparison with Howe's Approximation for Capacitance of a Rectangle

Howe[8] has shown that the capacitance of a plane conductor is well approximated by a value Q/P_{avg}, where the charge Q is assumed as uniformly distributed over the conductor and P_{avg} is the average value of the resulting potential distribution over the body. Proceeding thus, he found for a rectangle of height $2h$ and width $2w$ that

$$C=2(wh)^{1/2}/2\alpha^{1/2}[(\sinh^{-1}\alpha)/\alpha+\sinh^{-1}(1/\alpha)+\alpha/3+1/3\alpha^2-(\alpha^2+1)^{3/2}/3\alpha^2]$$
$$=2K(wh)^{1/2} \tag{8}$$

where $\alpha=w/h$ and corresponding values of α and K are as shown in Table III.

For the capacitance of a square, Howe's equation yields $C\cong 0.374$ μμf. The nearly exact value, as comprised in Fig. 4 and Table II, is 0.402 μμf. Thus Howe's value for the square is low. On the other hand, for the capacitance of a 3 by 1 rectangle, Howe's equation yields $C=2K(w/h)^{1/2}=3(0.36)=0.623$ esu $=0.692$ μμf, which value is in good agreement with the subarea value of $C=0.643$ esu $=0.714$ μμf. Here, the error in Howe's value is approximately one-half the error for the case of the square. Again, for the capacitance of a 20 by 1 cm rectangle, Howe's equation yields $C=20(0.532)=2.37$ esu $=2.64$ μμf which value is in excellent agreement with the value of 2.40 esu $=2.66$ μμf, to be extrapolated from the curve calculated by subareas and by comparison with the boundary curves of Fig. 6. Thus, it is to be inferred that for $\alpha\geqslant 3$, Howe's values are in substantial agreement with the exact values of the capacitance of a rectangle.

Accordingly, on the basis of this discussion, it follows that the curve of Fig. 7, plotted with the values of Table IV, yields an excellent approximation to the exact value of the capacitance of a rectangle of any ratio of w/h.

Summary

1. The basic theory of approximate calculation, by the use of subareas, of the capacitance of a rectangular plane area, and of the distribution of charge density over it, is concisely outlined.

2. Illustratively, the method of subareas is employed to obtain an accurate value for the capacitance of a 3 by 1 rectangle.

3. Comparison of Figs. 3 and 4 indicates that the final value of Fig. 3 may be inferred as very close to the corresponding exact value.

4. The normalized curve of Fig. 7 yields the approximate capacitance of a rectangle of any specified ratio w/h of width to length.

Appendix I

Assume that the rectangle of Fig. 8 has a uniform surface charge density of value q. The total charge on an incremental area $dxdy$ is $qdxdy$. Each increment of charge $qdxdy$ contributes to the potential at the center 0 an increment of potential $dV=qdxdy/r$. Accordingly, the total potential at point 0 is

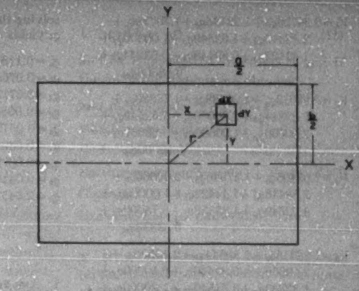


Fig. 8. Co-ordinate system for calculation of potential at center of uniformly charged rectangular subarea

$$V=4q\int_{y=0}^{\frac{b}{2}}\int_{x=0}^{\frac{a}{2}}\frac{dydx}{r}=4q\int_0^{\frac{b}{2}}\int_0^{\frac{a}{2}}\times\frac{dydx}{(x^2+y^2)^{1/2}}$$
$$=4q\left\{\frac{b}{2}\ln\left[\frac{a/2}{b/2}+\left(\frac{(a/2)^2}{(b/2)^2}+1\right)^{1/2}\right]+\frac{a}{2}\ln\left[\frac{b/2}{a/2}+\left(\frac{(b/2)^2}{(a/2)^2}+1\right)^{1/2}\right]\right\} \tag{9}$$

For the case of a 3 by 1 rectangle, plotted as the first approximation of Fig. 3,

$$V=4q\left\{\frac{1}{2}\ln\left[\frac{1.5}{0.5}+(3^2+1)^{1/2}\right]+\frac{3}{2}\ln\left[\frac{1}{3}+\left(\frac{1}{3^2}+1\right)^{1/2}\right]\right\}$$
$$=4q\{1.399\}=5.596q \tag{10}$$

Hence

$$C=Q_t/V=3q/5.596q=0.536\text{ esu}=0.597\ \mu\mu f \tag{11}$$

Appendix II

The set of equations for the charge densities, assumed uniform, of the subareas of Fig. 2 are

$$V_a=4.27175q_a+1.72741q_b+1.17833q_c+0.099636q_d+0.47360q_e+0.24253q_f+0.51233q_g+0.61159q_h+0.84853q_k+1.12403q_m$$

$$V_b=1.72741q_a+4.35027q_b+1.83289q_c+1.32715q_d+0.61067q_e+0.31622q_f+0.68973q_g+0.90322q_h+1.16439q_k+0.84853q_m$$

$$V_c=1.17833q_a+1.83289q_b+4.49909q_c+2.05787q_d+0.85355q_e+0.44720q_f+1.02333q_g+1.24253q_h+0.90322q_k+0.61159q_m$$

$$V_d=0.99636q_a+1.32715q_b+2.05787q_c+4.87904q_d+1.44720q_e+0.70711q_f+1.44720q_g+1.02333q_h+0.68973q_k+0.51233q_m$$

$$V_e=0.94720q_a+1.22135q_b+1.70710q_c+2.89440q_d+4.02549q_e+1.00000q_f+1.41422q_g+0.89440q_h+0.63244q_k+0.48506q_m$$

$$V_f=0.97012q_a+1.26488q_b+1.78880q_c+2.82844q_d+2.00000q_e+3.52549q_f+2.00000q_g+1.00000q_h+0.66667q_k+0.50000q_m$$

$$V_g=1.02466q_a+1.37946q_b+2.04666q_c+2.89440q_d+1.41422q_e+1.00000q_f+4.02549q_g+1.33333q_h+0.75000q_k+0.53333q_m$$

$$V_h=1.22318q_a+1.80644q_b+2.48506q_c+2.04666q_d+0.89440q_e+0.50000q_f+1.33333q_g+3.77549q_h+1.20000q_k+0.66667q_m$$

$$V_k=1.69706q_a+2.32878q_b+1.80644q_c+1.37946q_d+0.63244q_e+0.33333q_f+0.75000q_g+1.20000q_h+3.69216q_k+1.14285q_m$$

$$V_m=2.24806q_a+1.69706q_b+1.22318q_c+1.02466q_d+0.48506q_e+0.25000q_f+0.53333q_g+0.66667q_h+1.14285q_k+3.65049q_m \quad (12)$$

Imposing the condition $V_0=V_a=V_b=V_c=V_d=V_e=V_f=V_g=V_h=V_k=V_m$ and solving the resulting set of equations for the q_i yields

$$\begin{aligned} q_a&=0.11549V_0\\ q_b&=0.07698V_0\\ q_c&=0.07121V_0\\ q_d&=0.06862V_0\\ q_e&=0.06790V_0\\ q_f&=0.03945V_0\\ q_g&=0.03992V_0\\ q_h&=0.04169V_0\\ q_k&=0.04574V_0\\ q_m&=0.08441V_0 \end{aligned} \quad (13)$$

References

1. The Flow of Viscous Fluids Round Plane Obstacles, **R. Roscoe.** *Philosophical Magazine*, London, England, series 7, vol. 60, 1949, pp. 338–51.

2. Calculation of the Resistance to Ground and of Capacitance, **H. B. Dwight.** *Journal of Mathematics and Physics*, Cambridge, Mass., vol. 10, 1930–31, pp. 50–74.

3. Grounding Grids for High-Voltage Stations, **E. T. B. Gross, B. V. Chitnis, L. J. Stratton.** *AIEE Transactions*, vol. 72, pt. III, Aug. 1953, pp. 799–810.

4. Grounding Grids for High-Voltage Stations. II—Resistance of Large Rectangular Plates, **Eric T. B. Gross, Richard B. Wise.** *Ibid.*, vol. 74, pt. III, Oct. 1955, pp. 801–09.

5. Calculation of the Capacitance of a Circular Annulus by the Method of Subareas, **Thomas James Higgins, Daniel Kinseth Reitan.** *Ibid.*, vol. 70, pt. I, 1951, pp. 926–33.

6. Calculation of the Capacitance of a Cube by the Method of Subareas, **Daniel K. Reitan, Thomas J. Higgins.** *Journal of Applied Physics*, New York, N. Y., vol. 22, Feb. 1951, pp. 223–26.

7. The Solution of Boundary-Value Problems in Engineering by the Method of Subareas, **Thomas J. Higgins.** *Report No. 1*, Engineering Experiment Station, University of Wisconsin, Madison, Wisc., Sept. 1954.

8. The Capacity of Rectangular Plates and a Suggested Formula for the Capacity of Aerials, **G. W. O. Howe.** *The Radio Review*, London, England, vol. 1, Oct. 1919–June 1920, pp. 710–14.

9. The Application of Relaxation Methods to the Solution of Differential Equations in Three Dimensions, **D. N. De G. Allen, S. C. R. Dennis.** *Quarterly Journal of Mechanics and Applied Mathematics*, London, England, vol. VI, pt. 1, 1953, p. 87.

10. A Treatise on Electricity and Magnetism, Vol. I (book), **J. C. Maxwell.** Oxford University Press, London, England, third edition, 1893, pp. 148–54.

11. Electrical Researches (book), **H. Cavendish,** edited by **J. C. Maxwell.** Cambridge University Press, Cambridge, England, 1879, pp. 426–27.

12. Isoperimetric Inequalities in Mathematical Physics (book), **G. Polya, G. Szegö.** Princeton University Press, Princeton, N. J., 1951, pp. 78 and 157.

13. Tables of Integrals and Other Mathematical Data (book), **H. B. Dwight.** The Macmillan Company, New York, N. Y., 1947.

◆

Discussion

James D. Horgan (Marquette University, Milwaukee, Wisc.): In connection with boundary value problems, it is interesting to note that, in recent years, emphasis in analysis and research has shifted from problems involving Laplace's equation to problems involving the wave equation. Countless papers have appeared recently which analyze diffraction effects due to conductors of simple geometry. Yet there are still many unsolved problems which require the solution of Laplace's equation, subject to mixed boundary conditions. Although the equation itself is simpler, the boundary conditions in these problems make them extremely difficult to solve. The problem solved in this paper is typical of these.

The writer has had occasion recently to seek out and compare various means of solving Laplacian field problems with mixed boundaries, and has found that in many instances the method of subareas is the best, indeed, the only practical means, of solution. The recent paper by Higgins and Black,[1] in which the question of the capacitance of strip transmission lines was resolved and previous results shown to be in error, illustrates this fact. With the increased use of electronic computing facilities, this method will find ever-widening application by researchers in the field.

Reference

1. Rigorous Determination of Parameters of Microstrip Transmission Lines, **K. G. Black, T. J. Higgins.** *Transactions*, Professional Group on Microwave Theory and Techniques, Institute of Radio Engineers, New York, N. Y., vol. MTT-3, no. 2, March 1955.

Daniel K. Reitan and **Thomas J. Higgins:** The authors appreciate Prof. Horgan's favorable comment, and agree with him that often the subarea method offers the best practical means of solution of a boundary value problem characterized by Laplace's equation.

In addition to Prof. Horgan's note of his reference 1, the authors would direct attention to the pertinent complementary reference 1, wherein the results obtained analytically in the paper are substantiated by those obtained by electrolytic tank measurements, and to reference 2, which encompasses an account of the whole of Dukes' investigations.

Finally, Prof. Horgan's remarks as to the utility of the subarea approach in conjunction with high-speed computers are well substantiated by the interesting studies effected in precisely this manner by Gross:[3,4] who conjoined knowledge of the subarea method and a high-speed digital calculator to obtain accurate determinations of charge distribution over grounding plates for high-voltage systems.

References

1. Characteristic Impedance of Air-Spaced Strip Transmission Line, **J. M. C. Dukes.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 43, July 1955, p. 876.

2. An Investigation Into Some Fundamental Properties of Strip Transmission Line with the Aid of the Electrolytic Tank, **J. M. C. Dukes.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 103, pt. B, May 1956, pp. 319–33.

3. See reference 4 of the paper.

4. Grounding Grids for High-Voltage Stations. III—Resistance of Rectangular Grids, **Eric T. B. Gross, Robert S. Hollitch.** *AIEE Transactions*, vol. 75, pt. III, Oct. 1956, pp. 926–35.

Electrical Units and Dimensions

LEO YOUNG
MEMBER AIEE

Synopsis: A method is developed for converting between the units and the equations in different systems of units. The appearance or nonappearance of 4π as a constant of proportionality in Maxwell's equations may be attributed to switching the unit of solid angle between sphere and steradian. Treatment of angle as a dimension leads to a convenient and systematic way of conversion between different systems.

In electrical engineering the meter-kilogram-second (mks) rationalized system of units is generally adopted. Sometimes, however, the engineer encounters another system, such as one of the centimeter-gram-second systems commonly used in physics. He may work out the formula and convert his final answer from tables. Frequently he prefers to derive the formula over again by using mks rationalized units and equations throughout. In this paper, a method and a nomenclature will be developed to include the six systems of electrical units in general use. By explicitly introducing angle and solid angle, which are at the root of the diversity of constants, it becomes possible to write Maxwell's equations in a "universal" notation. By appealing to angle and solid angle constants as dimensional quantities, it becomes possible to introduce them into any equation in a given system as the missing dimensions. When the numerical values of the angle constants in the various systems of units are known, the equation can then be converted into any desired system.

A physical quantity, in addition to its four dimensions of fource F, length L, time T, and charge Q, becomes associated with two "angle" dimensions, plane angle U and solid angle S. (Force rather than mass is chosen only for convenience in dimensional computations. Charge is chosen for the electrical dimension to avoid fractional indexes.) To compare units of different systems, write down the six dimensions of the unit, and take the ratio with the respective units of F, L, T, Q, S, and U substituted.

It is not proposed to discuss the philosophical or the operational aspect of angle dimensions. In presenting this method it is only claimed that: (1) it is self-consistent, (2) it gives a simple procedure for converting between different systems, and (3) angle dimensions are justified if only because they provide a nomenclature easy to remember and to apply.

Angle Dimensions

In the mks rationalized system

$$\operatorname{div} \mathbf{D} = \rho \tag{1}$$

In the mks unrationalized system

$$\operatorname{div} \mathbf{D} = 4\pi\rho \tag{2}$$

Applying this to a point charge Q, and taking the flux over a spherical surface with the charge at its center,

$$\text{flux of } \mathbf{D} \text{ over the sphere} = Q \tag{3}$$

in rationalized mks units, and

$$\text{flux of } \mathbf{D} \text{ over the sphere} = 4\pi Q \tag{4}$$

in unrationalized mks units. Equation 4 can also be written

$$\text{flux of } \mathbf{D} \text{ per steradian} = Q \tag{5}$$

while equation 3 could equally well be written

$$\text{flux of } \mathbf{D} \text{ per sphere} = Q \tag{6}$$

Equations 5 and 6 are the same except for the units of solid angle. Equations 1 and 2 can be combined into the form

$$S^{-1} \operatorname{div} \mathbf{D} = \rho \tag{7}$$

where $S=1$ in mks rationalized units, and $S=4\pi$ in mks unrationalized units. Equations 5 and 6 show that the units of solid angle in the two systems are the sphere and the steradian respectively, and that the dimensions of $\mathbf{D}$ are $[L^{-2}QS]$.

As "div" is the limit of flux over a small closed surface, so "curl" is the limit of "rotation" round a small loop. The reciprocal angle constant U^{-1} is thus associated with curl as S^{-1} is associated with div.

Maxwell's equations can now be written

$$\operatorname{div} \mathbf{D} = S\rho \tag{8}$$

$$\operatorname{div} \mathbf{B} = 0 \tag{9}$$

$$U^{-1} \operatorname{curl} \mathbf{H} = S\mathbf{j} + \dot{\mathbf{D}} \tag{10}$$

$$U^{-1} \operatorname{curl} \mathbf{E} = -\dot{\mathbf{B}} \tag{11}$$

and by choosing the constants S and U the following systems are obtained.

mks rationalized: $S=1$, $U=1$
mks unrationalized: $S=4\pi$, $U=1$
electrostatic units (esu): $S=4\pi$, $U=1$
electromagnetic units (emu): $S=4\pi$, $U=1$
Gaussian: $S=4\pi$, $U=1/\gamma$
Heaviside-Lorentz units (hlu): $S=1$, $U=1/\gamma$

γ is a dimensionless number equal (approximately) to 3×10^{10}. The units of angle in each system are

mks rationalized: sphere, turn
mks unrationalized: steradian, turn
esu: steradian, turn
emu: steradian, turn
Gaussian: steradian, γ turns
hlu: sphere, γ turns

Application of Angle Dimensions

The application of angle dimensions can best be seen from two simple examples. First, let it be required to find the ratio of the mks rationalized unit of magnetic field to the mks unrationalized unit. Taking charge Q as the electrical dimension, the full dimensions (see Table I) of magnetic field are $[L^{-1}T^{-1}QSU]$. The required ratio is therefore as sphere is to steradian (since all the other units are the same), which is 4π.

Second, let it be given that the energy density of the electrostatic field in esu is $\mathbf{ED}/8\pi$. It is required to find the expression in mks rationalized units. The missing dimension is assumed to be S^x

$$[E][D][S^x] = [\text{energy}][L^{-3}]$$

and balancing dimensions (see Table I), $x=-1$. Now $S=4\pi$ in esu. Therefore, the expression in all systems of units is $(\mathbf{ED}/8\pi)\times(4\pi/S)$, which is $S^{-1}\mathbf{ED}/2$. In the mks rationalized system $S=1$, and so this reduces to $\mathbf{ED}/2$.

From Maxwell's equations 8, 9, 10, 11, and one force equation such as

$$\mathbf{E}Q = \mathbf{F} \tag{12}$$

the dimensions of any quantity can be found.

With charge Q as the electrical dimension, it is necessary to know that

$$\begin{aligned} 1 \text{ coulomb} &= \frac{1}{10} \text{ emu of charge} \\ &= 3\times 10^{9} \text{ esu (approximately)} \\ &= 3\times 10^{9} \text{ Gaussian units} \\ &= 3\sqrt{4\pi}\times 10^{9} \text{ hlu} \end{aligned}$$

in order to find the ratio of unit quantities in different systems of units. (Also, 1 newton $= 10^5$ dynes.)

Paper 56-984, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted April 2, 1956; made available for printing August 3, 1956.

Leo Young is with the Westinghouse Electric Corporation, Baltimore, Md.

To prove the relation between the hlu of charge and the esu of charge,

$$[\epsilon_0]=[F^{-1}L^{-2}Q^2S]$$

therefore

$$[Q]=[\epsilon_0^{1/2}F^{1/2}LS^{-1/2}] \tag{13}$$

The hlu and the esu of E_0, F, L are the same. Therefore

$$\frac{1\text{ esu of charge}}{1\text{ hlu of charge}}=\left(\frac{1\text{ hlu of solid angle}}{1\text{ esu of solid angle}}\right)^{1/2}$$

$$=\left(\frac{1\text{ sphere}}{1\text{ steradian}}\right)^{1/2}$$

$$=\sqrt{4\pi}$$

Table I gives the dimensions of some common quantities. A list of some typical dimensionally complete equations is given in the Appendix.

Summary

It has been shown that insertion of a constant S into the equations describing electromagnetic phenomena makes it possible to unite the equations of four systems (both mks systems, esu, and emu) into a single form. *Since S is a constant of proportionality in the defining equations,* it may be treated as a fifth dimension and introduced wherever it is missing in an equation by balancing dimensions. Moreover, it was shown how S takes the numerical value 1 or 4π according to the unit of solid angle chosen.

It was found convenient and natural to identify the dimensional constant S with solid angle. Combining the two then produces "dimensions in solid angle." This gives not only a method, but also a nomenclature based on the nature of the problem.

Similarly, introducing (plane) angle dimension U allows the two remaining systems of units, Gaussian and hlu, to be included.

Table I. Dimensions and Name in mks Rationalized Units of Some Common Quantities

Quantity	Symbol	Dimensions F	L	T	Q	S	U	Name in mks Rationalized Units
Charge	Q				1			coulomb
Electric field intensity	$\mathbf{E}$	1			−1			volt per meter
Electric displacement	$\mathbf{D}$		−2		1	1		coulomb sphere per meter²
Permittivity	ϵ_0	−1	−2		2	1		farad sphere per meter
Polarization	$\mathbf{P}$		−2		1			coulomb per meter²
Current	I			−1	1			ampere
Magnetic field intensity	$\mathbf{H}$		−1	−1	1	1	1	ampere-turn sphere per meter
Magnetic induction	$\mathbf{B}$	1	−1	1	−1		−1	weber per meter²
Wave impedance	Z_w	1	1	1	−2	−1	−1	ohm per turn sphere
Permeability	μ_0	1		2	−2	−1	−2	henry per meter sphere turn²
Magnetic flux	Φ	1	1	1	−1		−1	weber
Pole strength (defined by $p\mathbf{H}=\mathbf{F}$)	p	1	1	1	−1	−1	−1	weber per sphere
Intensity of magnetization	J	1	−1	1	−1	−1	−1	weber per sphere meter²
Electric potential, emf	V	1	1		−1			volt
Resistance	R	1	1	1	−2			ohm
Capacitance	C	−1	−1		2			farad
Inductance	$\mathcal{L}$	1	1	2	−2			henry
Conductivity	σ	−1	−2	−1	2			mho per meter
Resistivity	r	1	2	1	−2			ohm meter
Magnetomotive force, mmf	h			−1	1	1	1	ampere-turn sphere
Reluctance	$\mathcal{R}$	−1	−1	−2	2	1	2	ampere-turn sphere per weber

Note especially the difference in dimensions between such quantities as **D** and **P**, I and h, **B** and J, R and Z_w, Φ and p.

Appendix

Some typical dimensionally complete equations are as follows:

Force between two charges Q_1 and Q_2 in free space and distance d apart,

$$F=\frac{S}{4\pi}\frac{Q_1Q_2}{\epsilon_0 d^2} \tag{14}$$

Force on current element $Id\mathbf{s}$ in magnetic induction field **B**,

$$d\mathbf{F}=UId\mathbf{s}\times\mathbf{B} \tag{15}$$

Force per unit length between two infinite straight wires carrying currents I_1 and I_2 in free space and distance d apart,

$$F=\frac{U^2S}{2\pi}\mu_0\frac{I_1I_2}{d} \tag{16}$$

Energy density of electromagnetic field

$$=\frac{1}{2}S^{-1}(\mathbf{ED}+\mathbf{HB}) \tag{17}$$

Electric displacement in a medium subject to an electric field **E**, which produces a polarization **P**,

$$\mathbf{D}=\epsilon_0\mathbf{E}+S\mathbf{P} \tag{18}$$

Capacitance of parallel plate capacitor of area A, plate separation t, assuming parallel lines of force,

$$C=\frac{\epsilon_0}{S}\frac{A}{t} \tag{19}$$

Self-inductance of long solenoid of cross-section A, length b, and a total of N turns,

$$\mathcal{L}=U^2S\mu_0N^2A/b \tag{20}$$

Magnetic moment $d\mathbf{M}$ of current I flowing round an area $d\mathbf{A}$,

$$d\mathbf{M}=U\mu_0Id\mathbf{A} \tag{21}$$

References

1. Generalized Electrical Formulas, V. P. Hessler, D. D. Robb. *Electrical Engineering,* vol. 70, Apr. 1951, pp. 332–36.

2. Units and Terminology, L. Young. *Journal,* Institution of Electrical Engineers, London, England, vol. 1, New Series, 1955, p. 652.

3. The Theory of Dimensions, L. Young. *Ibid.,* vol. 2, New Series, 1956, p. 218.

◆

Discussion

F. T. Chapman (Bexleyheath, Kent, England): In this paper Mr. Young has shown the advantage of including dimensions for plane and solid angles in some of the dimensional equations used in electrical theory. Plane and solid angles are physical quantities but so long as they are defined as ratios, namely, arc per radius and area per (radius)² respectively, they are dimensionless; that is, there are no dimensional quantities the size of whose units can affect the size of the units of angle. But there are other methods of measuring plane angle in arbitrary units such as degrees, quadrants, or revolutions (turns), and, in the case of solid angle, there is the alternative of the sphere. The balance of equations will be affected and the measure numbers of quantities which are defined in terms including angular measurement will be changed by a change from one unit of angle to another. This consideration justifies the inclusion of a measure number for angle in equations 7–11, for example, and the inclusion of a dimension for angle in such dimensional expressions as Mr. Young gives for Q in equation 13.

Carl C. Chambers (University of Pennsylvania, Philadelphia, Pa.): This paper is a very material contribution to the field. In particular, it shows clearly that the 4π's that appear or do not appear in the various systems of units are identified with the concept of an angle. When the angle is defined as the ratio of two lengths without the specification of a constant, the logical unit of angle is steradian. Otherwise the units are said to lack coherency. Of course, in many systems of units the angle is defined as a proportion of a circle, or in the case of a solid angle as a proportion of a sphere. Under these conditions the terms, angle and solid angle, are related to the fundamental quantities of length in a quite different manner than in the case of their definitions which lead coherently to radians and steradians. In spite of its value in emphasizing the dependence of the presence or absence of 4π on the concept of angle, this paper in common with many other papers on this subject falls short of clarifying the confusing concepts of physical quantities, units, and dimensions. "Physical quantities" are measurable physical concepts. "Units" are samples of those measurable quantities. "Dimensions" are those characteristics of physical quantities which are specified by the algebraic powers of the chosen fundamental quantities which enter into the specification of the given physical quantities. Thus velocity has the dimensions, distance to the first power over time to the first power, and this is irrespective of the units of velocity or length or time. I am confident that Mr. Young did not mean this particular concept when he suggested that the S and the U in his paper be added dimensions. I might suggest that he substitute the words physical quantity and the meaning for his use of the word dimension in this respect. However, this would explicitly move into the middle of the real controversy which is going on in international circles concerning the meaning of rationalization. This controversy rests squarely with the interpretation of the definitions, first of physical quantity and second of unit. Some scientists considering the definition that physical quantities are those measurable concepts concerning an experimental setup, insist that this means that it is not a physical quantity until the definition in terms of the selected fundamental physical quantities is such that the reader of the definition knows explicitly how to measure that quantity.

I have suggested that this particular term should be "specific physical quantity." This would be in contrast to what might be called "generalized physical quantity" which would then mean a physical concept which was defined by a law of physics with a constant of proportionality yet to be determined. Thus in the paper, equation 3 which related the flux of **D** over a sphere as equal to Q would define **D** as a different specific physical quantity from his equation 4 which defines **D** by the statement flux of **D** over a sphere equal $4\pi Q$. According to my suggested definitions of two different physical quantities, the **D**'s in equations 3 and 4 would be different specific physical quantities while being the same generalized physical quantity. The difference between these two **D**'s as specific physical quantities is expressed explicitly in a different way in his equations 5 and 6.

This paper pedagogically presents a method which will help to show how the 4π's that appear or do not appear may be related back to their fundamental source, the concept of an angle. However, they do not contribute to the extent of solving the international contentions which are current with regard to the meaning of rationalization and its relationship to physical quantities on the one hand and to units on the other.

R. O. Kapp (Kennedy and Donkin, London, England): Mr. Young's proposal to raise angle to dimensional status is to be welcomed, for there are many occasions when it serves a useful purpose in dimensional equations. Science has suffered considerably in the past from the erroneous tradition that only three quantities, mass, length, and time, should be used. That tradition is not yet quite dead.

If I have any criticism of Mr. Young's method of relating rationalization to theory of dimensions, it is only that I could wish him to have gone further and made a more sweeping generalization. I have shown in a paper scheduled to be read before the Institution of Electrical Engineers (British) that the difference between rationalized and unrationalized expressions can be clarified if one gives dimensional status to a configuration measure. Any symmetrical area could then be expressed as σd^2, where d is the distance across the area. By a sound tradition this is equated to $\pi/4$ for a circle. If this convention is applied consistently, a system of units based on the 1-to-1 ratio is rationalized. If, however, the configuration factor is so chosen that its value is 1/16 for a circle, a system of units built on the 1-to-1 ratio is unrationalized. Table II, taken from my paper, makes this clear.

Table II. Alternative Systems of Units

Take as Basis the Definition of Field Strength $E=\psi/a\epsilon$, Where ψ/a Is the Flux Density

	Value		
	Conventional		Unconventional
Quantity	Coherent Rationalized	Uncoherent Unrationalized	Coherent Unrationalized
Relation between flux and charge	$\psi=Q$	$\psi=4\pi Q$*	$\psi=Q$
Area of circle	πr^2	πr^2	$r^2/4$†
Area of sphere	$4\pi r^2$	$4\pi r^2$	r^2†
Field strength E round a point charge	$Q/4\pi r^2\epsilon$	$Q/r^2\epsilon$	$Q/r^2\epsilon$

*Uncoherent.

†Unconventional.

An unrationalized system requires that one sacrifice either coherence or convention. In other words, an unrationalized system can only be coherent if expressions for areas are unrationalized.

J. J. Smith (General Electric Company, Schenectady, N. Y.): Mr. Young presents an interesting approach to the relations between the units in different electric systems by introducing the two units of solid angle, the sphere and the steradian. In the centimeter-gram-second (cgs) unrationalized system the unit used for a solid angle was the steradian. The mks unrationalized system can be similarly derived, using the steradian as a unit for the solid angle.

In the derivation of the rationalized mks system from current as the fourth fundamental unit no use is usually made of solid angles. It is based on the following definition of the ampere as given in the revision of ASA *C42* "Standard Definitions of Electrical Terms": "The ampere is a unit of current of such magnitude that if 1 ampere is maintained constant in each of two straight parallel conductors of infinite length and of negligible circular sections placed 1 meter apart in a vacuum, a force will be produced equal to 2×10^{-7} newtons per meter of length."

However, it is found that if solid angles are used, the sphere must be used as a unit in the rationalized system. Thus it might be expected that some of the formulas in a rationalized system will differ from those in an unrationalized system by some power of 4π which is the ratio of one sphere to one steradian.

His results appear to cover the six systems to which he refers. However, in its present form it has the disadvantage that it requires a change in the names of some of the units in the mksa system which have been worked out by standardizing bodies such as American Standards Association and International Electrotechnical Commission. For instance, in Table I the name of the unit for electric displacement density **D** is given as the "coulomb sphere per square meter" instead of the "coulomb per square meter" generally used.

This change in the name of accepted units may be avoided by giving a slightly different interpretation to the symbols S and U as shown in equation 22. Let us start with equation 3 which may be written

$$\int\int \mathbf{D}_r dA = Q \tag{22}$$

Following Maxwell we shall consider the symbols **D**, A, and Q to denote only numerical values. The subscript r denotes it is the numerical value of **D** in the rationalized system.

Equation 22 may be simplified for this discussion by assuming that Q is at the center of a sphere over which D is uniform. On integrating we get

$$4\pi R^2 D_r = Q \tag{23}$$

where R is the radius of the sphere and the suffix r is used to indicate that D_r is the numerical value in the rationalized system. In the following discussion of this equation we shall consider the cgs system both in the rationalized and unrationalized form. The reason for using the cgs system is that names have been given to the rationalized and unrationalized units in that system, namely the statcoulomb and the line. The unit corresponding to the statcoulomb is the coulomb in the mksa system but no name has been given to the unit which corresponds to the centimeter-gram-second unit "line" in the mksa system. Suggestions have been made that it be called the "pra-line."

We may associate numerical equation 23 with its appropriate unit by adding the unit statcoulomb to each side of the equation

$$4\pi R^2 D_r \text{ statcoulombs} = Q \text{ statcoulomb} \tag{24}$$

This gives

$$D_r \text{ statcoulombs per square centimeter} = \frac{Q}{4\pi R^2} \text{ statcoulombs per square centimeter} \tag{25}$$

The numerical equation corresponding to equation 23 in the unrationalized system is

$$R^2 D_u = Q \tag{26}$$

where D_u denotes the numerical value in the cgs unrationalized system. Since the unit in that system is the line we have

$$D_u \text{ lines per square centimeter} = \frac{Q}{R^2} \text{ lines per square centimeter} \tag{27}$$

In the unrationalized cgs system it was held that unit charge gave rise to 4π lines where the statcoulomb was the unit in the rationalized system and the line was the unit in the unrationalized system. Thus

$$1 \text{ statcoulomb} = 4\pi \text{ lines} \tag{28}$$

and

$$1 \text{ statcoulomb per square centimeter} = 4\pi \text{ lines per square centimeter} \tag{29}$$

But

$$\begin{aligned} 1\ \frac{\text{statcoulomb}}{\text{square centimeter}} &= 1\ \frac{\text{statcoulomb}}{\text{square centimeter}}\ \frac{4\pi \text{ steradians}}{\text{sphere}} \\ &= 4\pi\ \frac{\text{statcoulomb}}{\text{square centimeter}}\ \frac{\text{steradians}}{\text{sphere}} \end{aligned} \tag{30}$$

Thus, comparing this with equation 29 we may write

$$1\ \frac{\text{statcoulomb}}{\text{square centimeter}}\ \frac{\text{steradian}}{\text{sphere}} = 1 \text{ line per square centimeter} \tag{31}$$

Hence we obtain the unit of electric displacement density in the unrationalized system from the unit in the same rationalized system by multiplying it by the ratio of the steradian to the sphere.

This ratio of the steradian to the sphere is the same as Young's S^{-1} so we may write

$$S_u^{-1} = \frac{1 \text{ steradian}}{1 \text{ sphere}} = \frac{1}{4\pi} \tag{32}$$

where the subscript u denotes it is the value used in the case of an unrationalized system. Similarly we may write

$$S_r^{-1} = \frac{1 \text{ sphere}}{1 \text{ sphere}} = 1 \tag{33}$$

where the symbol r denotes it is the value used in the case of a rationalized system.

In the same way it can be shown in the case of magnetic field strength H that the rationalized unit, the ampere-turn per meter, is related to the unrationalized unit, the millioersted, by the equation,

$$1\ \frac{\text{ampere-turn}}{\text{meter}}\ \frac{\text{steradian}}{\text{sphere}} = 1 \text{ millioersted} \tag{34}$$

or

$$1 \text{ ampere-turn per meter} = S_u \text{ millioersteds} = 4\pi \text{ millioersteds} \tag{35}$$

The reason for the introduction of the factor 3×10^{10} in the Gaussian and Heaviside-Lorentz systems was that the electrical quantities were measured in the esu system and the magnetic quantities were measured in the emu system. See for instance Jeans "Electricity and Magnetism."[1] To convert the numerical value of the voltage in esu system to its numerical value in the emu system the following equation was used

$$1 \text{ stat volt} = 3 \times 10^{10} \text{ abvolts} \tag{36}$$

Thus, starting with the mks system equation 11 in the paper,

$$\text{curl } \mathbf{E} = -\dot{\mathbf{B}} \tag{37}$$

Integrating this around a complete circuit gives

$$e = \oint \mathbf{E}.ds = -\dot{\phi} \tag{38}$$

adding the units to this it becomes in the mks system

$$e \text{ volts} = -\dot{\phi} \text{ volts} \tag{39}$$

In the cgs emu system it is

$$e \text{ abvolts} = -\dot{\phi} \text{ abvolts} \tag{40}$$

and in the cgs esu system it is

$$e \text{ statvolts} = -\dot{\phi} \text{ statvolts} \tag{41}$$

The Gaussian and Heaviside-Lorentz systems used the equation in a form where e was the numerical value of the voltage in the esu system, and $\dot{\phi}$ was the numerical value of the voltage in the emu system. Thus inserting the units in this manner

$$\begin{aligned} e_s \text{ statvolts} &= -\dot{\phi}_m \text{ abvolts} \\ &= -\dot{\phi}_m \text{ statvolts} \times \frac{1 \text{ abvolts}}{1 \text{ statvolt}} \\ &= -\frac{1}{3 \times 10^{10}}\, \dot{\phi}_m \text{ statvolts} \end{aligned} \tag{42}$$

where the subscript s denotes it is the numerical value in the esu system and the subscript m denotes the numerical value in the emu system. Since the units are the same on both sides they may be omitted and the numerical equation written in the usual form. The subscripts s and m are usually omitted at the same time

$$e = -\frac{1}{3 \times 10^{10}}\, \dot{\phi} \tag{43}$$

We can now identify Young's U as follows. In a mixed system where the numerical values are given in statvolts on one side of the equation, and are given in abvolts on the other side (mixed system)

$$U_m^{-1} = \frac{1 \text{ statvolt}}{1 \text{ abvolt}} = 3 \times 10^{10} \tag{44}$$

The corresponding value when the numerical values are given in the same units on both sides of the equation is (nonmixed system)

$$U_n^{-1} = \frac{1 \text{ statvolt}}{1 \text{ statvolt}} = \frac{1 \text{ abvolt}}{1 \text{ abvolt}} = 1 \tag{45}$$

It may be helpful to write Young's equations 8 and 10 with the appropriate units on each side using the cgs system since names for all the units are available

$$\begin{aligned} \text{div } \mathbf{D}_r \text{ statcoulombs per cubic centimeter} &= \rho \text{ statcoulombs per cubic centimeter} \\ \text{div } \mathbf{D}_u \text{ lines per cubic centimeter} &= 4\pi\rho \text{ lines per cubic centimeter} \\ \text{curl } \mathbf{H}_r\ ab\ At \text{ per square centimeter} &= (i_m + \dot{D}_m)ab\ At \text{ per square centimeter} \\ &= \frac{1}{3 \times 10^{10}}(i_s + \dot{D}_s)\ ab\ At \text{ per square centimeter} \\ \text{curl } \mathbf{H}_u \text{ gilberts per square centimeter} &= 4\pi(i_m + \dot{D}_m) \text{ gilberts per square centimeter} \\ &= \frac{4\pi}{3 \times 10^{10}}(i_s + \dot{D}_s) \text{ gilberts per square centimeter} \end{aligned} \tag{46}$$

Here the subscripts r, u, m, and s are used for the sake of clarity. It should be recognized that when the units are dropped from each side of the equation to obtain numerical equations, the subscripts are also usually dropped, giving the following set of numerical equations for curl **H**

$$\begin{aligned} \text{curl } \mathbf{H} &= i + \dot{\mathbf{D}} \\ \text{curl } \mathbf{H} &= \frac{1}{c}(i + \dot{\mathbf{D}}) \\ \text{curl } \mathbf{H} &= 4\pi(i + \dot{\mathbf{D}}) \\ \text{curl } \mathbf{H} &= \frac{4\pi}{c}(i + \dot{\mathbf{D}}) \end{aligned} \tag{47}$$

Each of these numerical equations belongs to a different system. The system to which each belongs is readily found from equation 10, or from equation 46, if only the cgs system is considered.

By interpreting Young's S and U in the manner shown here, the change in the names of the mks units given in Table I is avoided. This seems quite desirable since these names have been worked out in the American Standard Association and International Electrochemical Commission and are used extensively by many authors.

REFERENCE

1. ELECTRICITY AND MAGNETISM (book), J. H. Jeans. Cambridge University Press, London, England, second edition, 1911, p. 500.

Leo Young: It is a pleasure to note in the discussions so much fundamental agreement with the method described. This approach has evidently been shared by others who have concerned themselves with the subject.

Dr. Chapman points out that the balance of equations is affected and the measure numbers of quantities change, which are both the result of changing units of angle. So long as angle is defined as a ratio, with always a fixed constant of proportionality, it remains dimensionless and no way exists to handle a situation where a change of angle units occurs.

Dr. Chambers distinguishes between physical quantities, units, and dimensions. In trying to be brief, and wishing to emphasize the main argument, I did not refer to them. To take Dr. Chambers' example, velocity is defined as length divided by time. Let us introduce a constant of proportionality, k, and write

$$\text{velocity} = k \times \frac{\text{length}}{\text{time}} \tag{48}$$

Then the constant k would appear wherever a physical quantity is defined directly in terms of velocity, rather than in terms of length and time separately.

Every physical quantity has to be defined by one of two alternative methods: (1) in terms of an arbitrary unit by direct comparison, or (2) in terms of other quantities by some arbitrary process. If the former, it is referred to as a fundamental (or primary) quantity; if the latter, as a derived (or secondary) quantity. For a derived quantity the definition takes the form of an equation such as equation 48, and involves an arbitrary constant of proportionality. It is largely a matter of choice whether we decide to take a given physical quantity as fundamental or derived. However, there is a minimum number of fundamental quantities less than which we cannot have. In mechanics there are three, in electromagnetism there is also a fourth fundamental quantity.

The unit of a derived quantity changes whenever we change the units of the fundamental quantities on which it depends through its defining equation, or whenever we change the constant of proportionality in that equation.

The equations do not change with the fundamental units, but only with the constants or proportionality.

With any electrical quantity there are associated four primary dimensions which may be taken to be force F, length L, time T, and charge Q. These dimensions are assigned to a quantity through its defining equation. The process may be summarized as follows. Strip the defining equation of its constant of proportionality and all operator symbols (such as differential or integral operators), leaving only the symbols F, L, T, and Q. If necessary, first reduce all previously defined derived quantities to their primary dimensions in F, L, T, and Q. This turns the defining equation into the dimensional formula of the quantity defined. It tells how the unit of the derived quantity changes when the units of the fundamental quantities are changed.

In the process of obtaining the dimensional formula, the information carried by the constant of proportionality in the defining equation has been lost. For most equations and in most systems this constant is unity, and the information is of no consequence. Sometimes, however, this information is worth preserving.

Consider for instance equation 1 which defines D. When div is replaced by its dimensional value, and S is dropped, it gives the dimensional formula of D,

$$[D]=[L^{-2}Q] \qquad (49)$$

Suppose however S is left in, so that

$$[D]=[L^{-2}QS] \qquad (50)$$

where $[S]$ is kept as a fifth dimension. We speak of $[S]$ as a secondary dimension, to distinguish it from the minimum number of (four) primary dimensions. Its usefulness lies in the possibility of treating a change of equations as a change of units, since the generalized equations remain the same.

The constant k in equation 48 could be left in the dimensional formula for velocity, and would be referred to as a secondary dimension of velocity in the sense just explained. However, k, S, and U do not possess primary dimensions in the sense of F (or M), L, T, and Q.

I think the foregoing statement is akin to the contribution from Professor Kapp, who makes a similar point.

Dr. Smith gives an interesting derivation of certain relations between units of different systems, using the two values of the constant S: $S_u=4\pi$ and $S_r=1$. He goes on to show how the two values of U, $U_m \doteqdot 1/(3\times10^{10})$ and $U_u=1$, are the result of mixing or not mixing esu with emu (U_m is the same as $1/\gamma$. Dr. Smith's c in equation 47 is my γ. I wished to avoid using c for U_m which is a dimensionless number $\gamma \doteqdot 3\times10^{10}$, while c is usually taken for the velocity of light, which has this particular numerical value only when it is measured in centimeters per second. Thus in my notation c is not γ, but $c=\gamma$ centimeters per second = 186.000 miles per second, etc.).

He thus obtains the same results as using angle dimensions. He has the advantage of being thoroughly familiar with the basic relationships between units, whereas my method is intended for those who are not so familiar, or who prefer a quicker method, which has most of this information already "built into it." Once understood, it requires the minimum of effort when applied to practical problems.

I did not suggest that the names of the "generalized" or "transitional" units in Table I be adopted by the ASA or IEC. They should be thought of as units in transition between two systems. On many occasions this nomenclature would be helpful, particularly when parties using different systems wish to communicate. I also hope that this approach will help engineers and scientists to a better understanding of the subject. How widely it is to be used depends on them.

A New Method for Treating Nonlinear Problems with Applications to Iron-Cored Oscillatory Circuits

I—Subharmonics Do Not Exist

M. KAMAL GOHAR
NONMEMBER AIEE

Synopsis: Nonlinear circuits are now widely used in many control regulation systems or other industrial applications. Most of the solutions of such circuits have been based on either the step-by-step or successive approximation methods which do not always lead to the correct results, except in some special cases. Also, such calculations are complex so that generally the first approximation only is adopted. In this paper a new method for attacking nonlinear problems is suggested. By a certain representation of the nonlinear given relation the amount of work is greatly reduced and the calculations can even be carried out by an ordinary slide rule. According to the author's knowledge, the solution for the case of iron-cored oscillatory circuits and similar cases when the supply frequency, current or voltage, contains considerable harmonics is obtained for the first time. With respect to hysteresis action a simplified solution is given. Experimental results are in good agreement with the theoretical solutions. Harmonics up to the fifth are the only ones used herein to represent the required results. Such representations can be carried out with any number of harmonics by similar process.

PRACTICALLY all single-valued nonlinear relations can be satisfactorily represented by two or three straight lines. Fig. 1 shows the representation of a saturation curve of iron as suggested by Boyajian.[1] In his solution, he used two linear relations, one for each region, and remarked that the transition back and forth between them introduced a transient term into the equation of each region. However, his method led to complicated simultaneous equations which could not be solved except for some special cases. Other authors preferred to express the nonlinear relation in other forms such as Fourier or polynomial series. Figs. 2 and 3 show such a representation by Fourier series for a plate-current–grid-voltage characteristic of a triode and the current-flux relation for an iron-cored coil, respectively.[2,3] Fig. 4 shows another representation for a B-H curve by the use of polynomial series.[4] It took hours to make such representations. Most of the solutions using the foregoing representa-

---

Paper 56-983, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1-5, 1956. Manuscript submitted June 10, 1955; made available for printing August 31, 1956.

M. KAMAL GOHAR is with Cairo University, Cairo, Egypt.

The author is indebted to Dr. Abdel Samie Moustafa for his valuable advice, and to Mukhtar Nashid for his permission to use some of his experimental work.

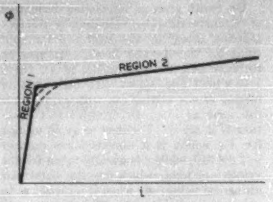
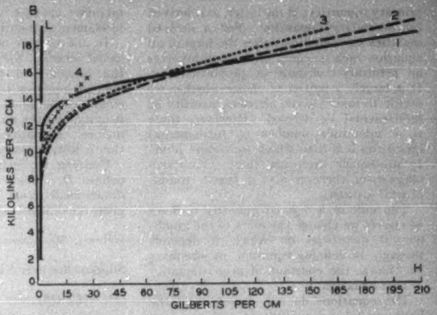


Fig. 1. Representation of saturation curve by two straight lines

Fig. 4. Representation of B-H curve by polynomial series



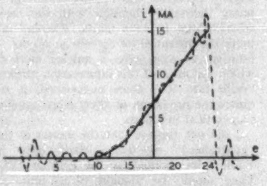


Fig. 2. Representation of i-e curve by Fourier series

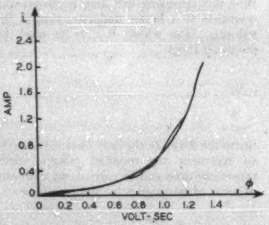


Fig. 3. Representation of i-φ curve by Fourier series

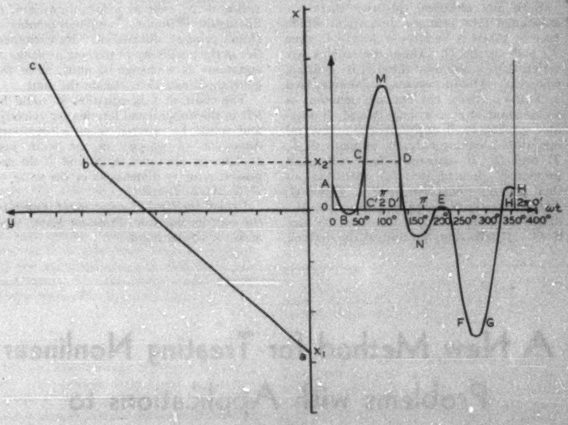


Fig. 5. Oscillatory function under the influence of a nonlinear relation represented by two straight lines

tions were based on either the step-by-step or successive approximation methods which do not always lead to correct results and so they were applied only to some special cases or to small nonlinear systems. In this paper nonlinear relations are represented by certain straight lines in such a way as to avoid successive approximations.

Consider Fig. 5, in which the relation between y and x can be represented by the straight lines ab and bc. Mathematically these lines can be given by

For ab: $y=m_1x_1+C_1,\ x_1<x<x_2$ (1)

For bc: $y=m_2x+C_2,\ x_2<x$ (2)

If x varies periodically with time t and is of the form

$$x=F(\omega t)=A_0+\sum_{n=1} A_n\cos n\omega t+\sum_{n=1} B_n\sin n\omega t \quad (3)$$

represented by the curve $ABCMDNEFGH$, the corresponding expression for y will depend on whether $x_1<x<x_2$ or $x_2<x$. If x is in the first range, $y=m_1x+c_1$, that is,

$$y=m_1[A_0+\Sigma A_n\cos n\omega t+\Sigma B_n\sin n\omega t]+C_1 \quad (4)$$

In general let θ_1 and θ_2 be the values of ωt at which $x=x_2$ represented by the points C and D in Fig. 5 In most practical cases two values of ωt are considered. Dividing $ABCMDNEFGH$ into two parts, $ABCC'D'DEFGH$ can be represented by a Fourier series $F_1(\omega t)$ and therefore $y=m_1F_1(\omega t)+c_1$, and $OC'CMDD'H'$ which can be represented by another Fourier series $F_2(\omega t)$, and hence $y=m_2F_2(\omega t)+C_2$. Therefore, the general expression for y is

$$y=m_1F_1(\omega t)+C_1(\omega t)+m_2F_2(\omega t)+C_2(\omega t) \quad (5)$$

where $C_1(\omega t)$ is a Fourier series to represent a curve of constant value C_1 from $\omega t=0$ to $\omega t=\theta_1$ and from $\omega t=\theta_2$ to $\omega t=2\pi$ while between θ_1 and θ_2 it has a zero value. Similarly, $C_2(\omega t)$ represents a curve of constant value C_2 between $\omega t=\theta_1$ and $\omega t=\theta_2$ and has a zero value for all other values of ωt from $\omega t=0$ to $\omega t=2\pi$.

Sometimes it may be more suitable to write equation 5 in the form:

$$y=m_1F(\omega t)+(m_2-m_1)F_2(\omega t)+(C_2-C_1)\times(\omega t)+C_1 \quad (6)$$

where $(C_2-C_1)(\omega t)$ represents a curve of constant value C_2-C_1 from $\omega t=\theta_1$ to $\omega t=\theta_2$ and has a zero value for all other values of ωt from $\omega t=0$ to $\omega t=2\pi$.

In order to reduce the amount of work, Fourier series for each of $y=A_m\cos m\omega t$, $y=B_m\sin m\omega t$, $y=C_m$ from $\omega t=\theta_1$ to

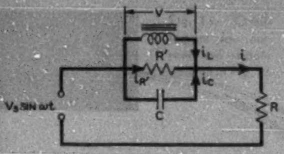


Fig. 6. A nonlinear oscillatory circuit

$\omega t=\theta_2$ with $y=0$ for all other values of ωt from $\omega t=0$ to $\omega t=2\pi$ are given in Appendix I.

The advantage of this method is that after determining the limits of ωt such as θ_1 and θ_2, the theory of superposition can be used. It can be used in two ways with advantage in the solution of differential equations for electric circuits (or otherwise) containing a nonlinear element. Although the first one is longer, not as reliable, and cannot generally be adopted except for small nonlinear cases, it is introduced here as it gives the fundamental idea of the solution.

Theoretical Analysis

A nonlinear differential equation which appears in many applications, the oscillatory circuit shown in Fig. 6, for example, is

$$A\frac{d^2\phi}{dt^2}+B\frac{d\phi}{dt}+i_L=K\sin\omega t \tag{7}$$

The relation between ϕ and i_L is nonlinear and skew. For positive values of ϕ and i_L the relation is shown and represented in Fig. 7. Let ϕ_1 and ϕ_2 be the values of ϕ at the end of the straight-line parts. As a first approximation assuming only the first-line part oa of the $\phi-i_L$ relation, the solution of the differential equation 7 by the ordinary well-known method gives

$$\phi=\phi_m\sin(\omega t+\psi) \tag{8}$$

If $\phi_m>\phi_2$, ϕ can be expressed as in equation 3 and i_L as in equation 6. (In this case θ_1 and θ_2 are equal to the values of ωt which make equation $8=\phi_1$ while, due to the third line bc, θ_3 and, θ_4 should be similarly introduced, thus only equation 8 will be equal to ϕ_2.) From the nature of this and similar problems, only odd harmonics are considered. Substituting the expressions for ϕ and i_L of equations 3 and 6 in equation 7 and equating the coefficients of similar terms in both sides will give linear equations in A_1, B_1, A_3, B_3, etc., which can be solved by well-known methods leading to

$$\phi=A_1\cos\omega t+A_3\cos 3\omega t+A_5\cos 5\omega t+ \\ B_1\sin\omega t+B_3\sin 3\omega t+B_5\sin 5\omega t \tag{9}$$

Table I. Approximations of Coefficients

Coefficient	Second Approximation	Third Approximation
A_1	−1.111	−1.13
B_1	−0.144	−0.159
A_3	−0.0384	−0.0389
B_3	−0.01	−0.19
A_5	0.00057	−0.0011
B_5	−0.00157	−0.0028

Harmonics up to the fifth were found to be generally sufficient.

As a second approximation, θ_1, θ_2, θ_3, and θ_4 are recalculated from equation 9 instead of equation 8, and the corresponding values of A_1, B_1, A_3 . . . etc., can be deduced by a similar procedure. This process should be repeated until two successive values for θ_1, θ_2, θ_3, and θ_4 (or A_1, B_1, A_3, etc.) will be nearly unchanged.

By substituting 12×10^{-6},1/390, and 0.746 for A, B, and K in equation 7, Table I gives the second and third approximations of A_1, B_1, A_3, B_3, and A_5 and B_5.

Consider the third approximation

$$V=\frac{d\phi}{dt}=289\sin(\omega t-8°)-0.113\times \\ \cos(3\omega t+62°36')-0.0133\times \\ \cos(5\omega t+22°\ 42) \tag{10}$$

This example has been solved analytically and checked experimentally by Mostafa and Shishini.[8] Analytically they found

$$V=290\sin(\omega t-3°20')+0.104\cos(3\omega t- \\ 82°)+0.0126\cos(5\omega t-96°46') \tag{11}$$

Fig. 8 shows the experimental oscillograph which was given by them. In this illustration the supply voltage is also shown 180 degrees out of phase. Although equations 10 and 11 may differ in the coefficients, they have nearly the same shape as shown in Fig. 9. (Another example of this is seen from Fig. 4 where the B-H curve could be represented by different polynomial series with different coefficients of the similar terms.)

As the foregoing analysis is based on successive approximations, it will not always lead to correct results as previously mentioned, especially if the initial values are not close to the final required ones or considerable harmonics are produced. Therefore, there is a need for another method which does not depend on successive approximations.

In the example given, the values of θ_1 and θ_2 can be directly determined with reasonable accuracy by the use of another equivalent representation of the nonlinear given relation. The ϕ-i_L relation given in Fig. 7 will be represented by three straight lines, oa, ab, and bc, as shown in Fig. 10 in which oa is parallel to bc and ab is perpendicular to the i_L axis. In this way the term $(m_2-m_1)\,F_2(\omega t)$ of equation 6 will be cancelled. The result is that θ_1 and θ_2 can be determined without successive approximations as shown in Appendix II. (Sometimes more than one real value for θ_1 and θ_2 can be obtained. In this case more than one solution may exist; a phenomenon which was illustrated by many previous authors, e.g., Thomson[4] and Stoker.[6])

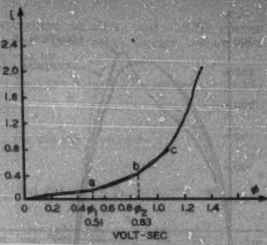


Fig. 7. Representation of I-φ curve in Fig. 3 by straight lines

With this second representation the solution of the previous example will be given by

$$v=258(\sin\omega t-0.3\cos\omega t+0.206\times \\ \sin 3\omega t+0.062\cos 3\omega t+0.000979\times \\ \sin 5\omega t-0.01121\cos 5\omega t) \tag{12}$$

Equation 12 is represented in Fig. 9 where it can be seen that it gives much better results. The second method of representation is similarly applied to the circuit shown in Fig. 11 (by separating the losses for the iron-cored coil the values of R' corresponding to the eddy-current losses can be determined). Neglecting the hysteresis effect for the time being, the differential equation will be

$$\left(1+\frac{R}{R'}\right)\frac{d^2\phi}{dt^2}+\frac{1}{CR'}\frac{d\phi}{dt}+R\frac{di_L}{dt}+\frac{i_L}{C} \\ =\omega V\cos\omega t \tag{13}$$

Fig. 8. Oscillogram for coil and supply voltages in Fig. 6. The latter is 180 degrees out of phase

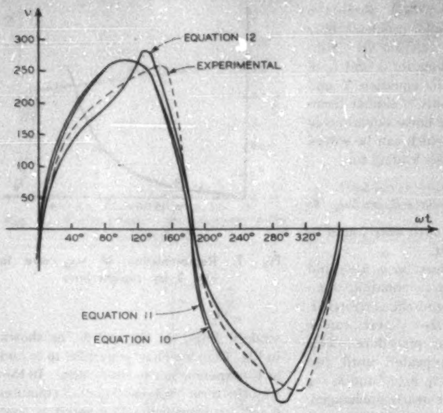


Fig. 9. Comparison between experimental and theoretical results for Fig. 6

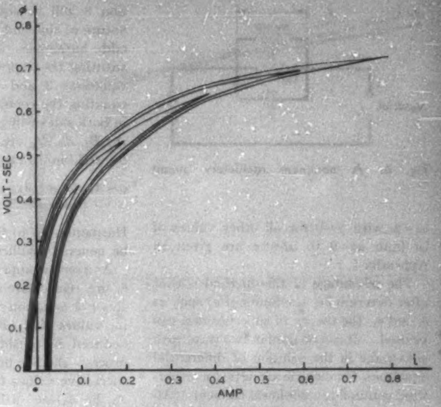


Fig. 12. φ-i loop curves for iron coil in Fig. 11

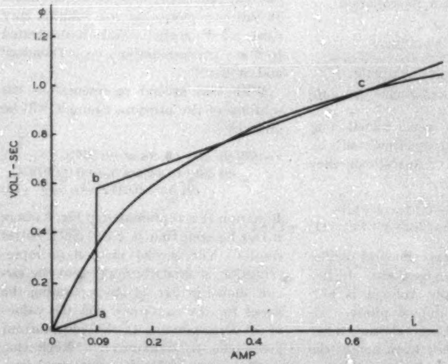


Fig. 10. Suggested method for representing φ-i curve in Fig. 3

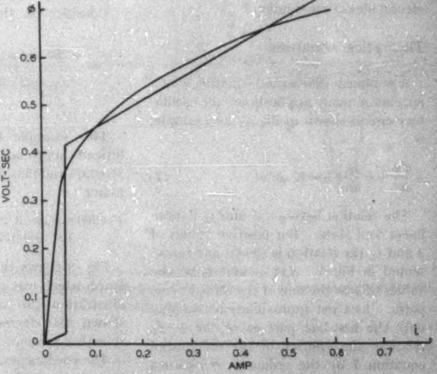


Fig. 13. Suggested representation of φ-i curve of Fig. 12

Working on the single-valued relation between ϕ and i_L which passes by the tips of the hysteresis loops (Fig. 12 shows the upper parts of such loops), the suggested representation would be as shown in Fig. 13 and it can be found that the following obtains

$$i=i_{R'}+i_L=0.29725\cos\omega t+0.0852\times\cos 3\omega t-0.09316\cos 5\omega t+0.1196\sin\omega t+0.19234\sin 3\omega t+0.05264\sin 5\omega t \quad (14)$$

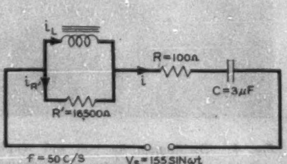


Fig. 11. A nonlinear oscillatory circuit

From equation 14 the rms value of $i=i_{\text{rms}}=0.282$ ampere. Experimentally $i_{\text{rms}}\doteq 0.26$ ampere, while its wave form is as shown in Fig. 14, where the supply voltage is also represented. From Fig. 15 it can be seen that equation 14 is in good agreement with the experimental results. In this example if it is required to take the hysteresis effect into consideration, the maximum value of i_L from equation 14 or Fig. 15 would equal to 0.62 ampere corresponding to $\phi_{\max}=0.693$ volt-second. From the area of the corresponding hysteresis loop and the frequency, the loss per second due to the hysteresis action would equal 3.865 watts which is equivalent to a resistance $R_{\text{hys}}=12{,}100$ ohms in shunt with R' (assuming the voltage on the iron-cored coil is not changed). This will lead to

$$i=0.2831\cos\omega t-0.00622\cos 3\omega t-0.09735\cos 5\omega t+0.1816\sin\omega t+0.2069\sin 3\omega t-0.0227\sin 5\omega t \quad (15)$$

Equation 15 is also represented in Fig. 15.

Fig. 14. Oscillogram for current and supply voltage in Fig. 11

The harmonics obtained in equation 14 or 15 are greater than 33 per cent. A similar result can be obtained if the supply voltage contains considerable harmonics. According to the author's knowledge, this is the only general solution found so far which can be used satisfactorily in the cases of such percentage of harmonics in the supply.

Conclusions

The suggested representation of the nonlinear single-valued relation is quicker than the other methods used before. It can be used to solve all types of nonlinear electric circuits, and similar cases, which contain a nonlinear element with great advantages, especially if the required results contain considerable harmonics. As the solutions given by other methods are restricted, Dr. Barrow has pointed out the importance of finding such a general solution.[2]

Appendix I

According to Fig. 16, *ab* is a part of curve $A_m \cos m\omega t$ when $m=1$. By Fourier analysis the curve *oa'abb'c* (and the similar ones when m equals other integer values) can be given by

$$a_{oc}+\Sigma a_{nc}\cos n\omega t+\Sigma b_{nc}\sin n\omega t \tag{16}$$

where

$$a_{oc}=\frac{A_m}{2\pi}\,\frac{\sin m\theta_2-\sin m\theta_1}{m}$$

$$a_{nc}=\frac{A_m}{2\pi}\left[\frac{\sin(n+m)\theta_2}{n+m}+\frac{\sin(n-m)\theta_2}{n-m}-\frac{\sin(n+m)\theta_1}{n+m}-\frac{\sin(n-m)\theta_1}{n-m}\right]\text{if } n\neq m$$

$$a_{nc}=\frac{A_m}{2\pi}\left[\theta_2-\theta_1+\frac{\sin(n+m)\theta_2}{n+m}-\frac{\sin(n+m)\theta_1}{n+m}\right]\text{if } n=m$$

$$b_{n,}=\frac{A_m}{2\pi}\left[\frac{\cos(n+m)\theta_1}{n+m}+\frac{\cos(n-m)\theta_1}{n-m}-\frac{\cos(n+m)\theta_2}{n+m}-\frac{\cos(n-m)\theta_2}{n-m}\right]\text{if } n\neq m$$

$$=\frac{A_m}{2\pi}\left[\frac{\cos(n+m)\theta_1}{n+m}-\frac{\cos(n+m)\theta_2}{n+m}\right]\text{if } n=m$$

According to Fig. 17, *ab* is a part of the curve $B_m \sin m\omega t$ when $m=1$. By Fourier series curve *oa'abb'c*, and similar curves when m equals other integer values, can be given by

$$a_{os}+\Sigma a_{ns}\cos n_{\omega}t_{ns}+\Sigma b_{ns}\sin n\omega t \tag{17}$$

where

$$a_{os}=\frac{B_m}{2\pi}\,\frac{\cos m\theta_1-\cos m\theta_2}{m}$$

$$a_{ns}=\frac{B_m}{2\pi}\left[\frac{\cos(n-m)\theta_2}{n-m}-\frac{\cos(n-m)\theta_1}{n-m}-\frac{\cos(n+m)\theta_2}{n+m}+\frac{\cos(n+m)\theta_1}{n+m}\right]\text{if } n\neq m$$

$$a_{ns}=\frac{B_m}{2\pi}\left[\frac{\cos(n+m)\theta_1}{n+m}-\frac{\cos(n+m)\theta_2}{n+m}\right]\text{if } n=m$$

$$b_{ns}=\frac{B_m}{2\pi}\left[\frac{\sin(n-m)\theta_2}{n-m}-\frac{\sin(n+m)\theta_2}{n+m}-\frac{\sin(n-m)\theta_1}{n-m}+\frac{\sin(n+m)\theta_1}{n+m}\right]\text{if } n\neq m$$

$$=\frac{B_m}{2\pi}\left[\theta_2-\theta_1-\frac{\sin(n+m)\theta_2}{n+m}+\frac{\sin(n+m)\theta_1}{n+m}\right]\text{if } n=m$$

According to Fig. 18, *ab* is a part of a straight line parallel to the x axis and its ordinate $=C_m$. By Fourier analysis, curve *oa'abb'c* is given by

$$a_o+\Sigma a_n\cos n\omega t+\Sigma b_n\sin n\omega t \tag{18}$$

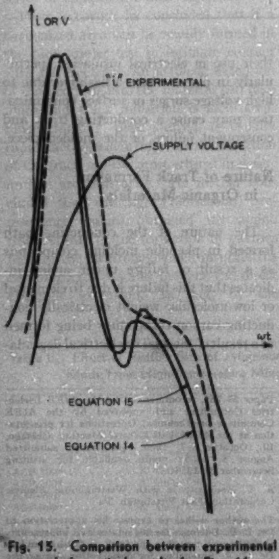


Fig. 15. Comparison between experimental and theoretical results for Fig. 11

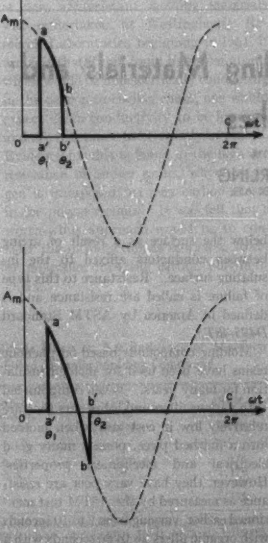


Fig. 16. A part of the curve $A_m \cos wt$

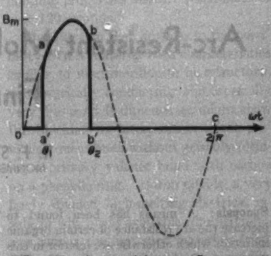


Fig. 17. A part of the curve $B_m \sin wt$

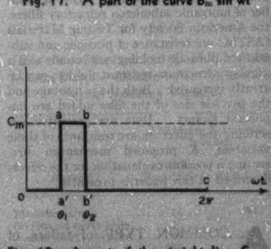


Fig. 18. A part of the straight line $C_m=$ constant

where

$$a_0=\frac{C_m(\theta_2-\theta_1)}{2\pi}$$

$$a_n=\frac{C_m}{\pi}\,\frac{\sin n\theta_2-\sin n\theta_1}{n}$$

$$b_n=\frac{C_m}{\pi}\,\frac{\cos n\theta_1-\cos n\theta_2}{n}$$

Appendix II

From Fig. 7, the $\phi-i_L$ relation can be expressed as follows:

$$\phi=0.764\,i_L \text{ from } i_L=-0.09 \text{ ampere to } i_L=0.09 \text{ ampere} \quad (19)$$

$$\left.\begin{aligned}&\phi=0.764\,i_L+0.515 \text{ when } i_L>0.09 \text{ ampere}\\ &\phi=0.764\,i_L-0.515 \text{ when } i_L<-0.09 \text{ ampere}\end{aligned}\right\} \quad (20)$$

Considering the harmonics up to the fifth only, i_L can be put in the form:

$$i_L=A_1\cos\omega t+A_3\cos 3\omega t+A_5\cos 5\omega t+B_1\sin\omega t+B_3\sin 3\omega t+B_5\sin 5\omega t \quad (21)$$

If θ_1 and θ_2 are the values of ωt which make i_L in equation 21 $=0.09$, then by the use of equations 3, 6, and Appendix I:

$$\begin{aligned}\phi=\;&0.764(A_1\cos\omega t+A_3\cos 3\omega t+A_5\cos 5\omega t+B_1\sin\omega t+B_3\sin 3\omega t+B_5\sin 5\omega t)+\\ &\frac{2\times 0.515}{\pi}\left\{(\sin\theta_2-\sin\theta_1)\cos\omega t+\frac{(\sin 3\theta_2-\sin 3\theta_1)}{3}\cos 3\omega t+\frac{(\sin 5\theta_2-\sin 5\theta_1)}{5}\cos 5\omega t+\right.\\ &(\cos\theta_1-\cos\theta_2)\sin\omega t+\frac{(\cos 3\theta_1-\cos 3\theta_2)}{3}\sin 3\omega t+\left.\frac{(\cos 5\theta_1-\cos 5\theta_2)}{5}\sin 5\omega t\right\}\end{aligned} \quad (22)$$

By substituting equations 21 and 22 in equation 7 and equating the coefficients of similar terms in both sides (A, B, and K are equal to 12×10^{-4}, 1/390, and 0.746, respectively) A_1, B_1, A_3, etc., will be given by

$$\left.\begin{aligned}A_1&=-0.885-0.000202(\sin\theta_2-\sin\theta_1)-0.51(\cos\theta_1-\cos\theta_2)\\ B_1&=0.775-0.000202(\cos\theta_1-\cos\theta_2)+0.51(\sin\theta_2-\sin\theta_1)\end{aligned}\right\} \quad (23)$$

$$\left.\begin{aligned}A_3&=-0.1721(\sin 3\theta_2-\sin 3\theta_1)-0.094(\cos 3\theta_1-\cos 3\theta_2)\\ B_3&=-0.1721(\cos 3\theta_1-\cos 3\theta_2)+0.094(\sin 3\theta_2-\sin 3\theta_1)\end{aligned}\right\} \quad (24)$$

$$\left.\begin{aligned}A_5&=-0.0939(\sin 5\theta_2-\sin 5\theta_1)-0.00121(\cos 5\theta_1-\cos 5\theta_2)\\ B_5&=-0.0939(\cos 5\theta_1-\cos 5\theta_2)+0.00121(\sin 5\theta_2-\sin 5\theta_1)\end{aligned}\right\} \quad (25)$$

From equations 21, 23, 24, and 25, considering the cases when $i_L=0.09$ ampere while $\omega t=\theta_1$ or θ_2, equations 26 and 27 can be deduced:

$$\begin{aligned}&1.16\sin(\theta_1-49.5^\circ)-0.000202\times\sin(\theta_2-\theta_1)+0.51\cos(\theta_2-\theta_1)-0.172\times\sin 3(\theta_2-\theta_1)+0.094\cos 3(\theta_2-\\ &\theta_1)-0.0939\sin 5(\theta_2-\theta_1)+0.0012\times\cos 5(\theta_2-\theta_1)-0.6952=0\end{aligned} \quad (26)$$

$$\begin{aligned}&1.16\sin(\theta_2-49.5^\circ)-0.000202\times\sin(\theta_2-\theta_1)-0.51\cos(\theta_2-\theta_1)-0.172\times\sin 3(\theta_2-\theta_1)-0.094\cos 3(\theta_2-\\ &\theta_1)-0.0939\sin 5(\theta_2-\theta_1)-0.0012\times\cos 5(\theta_2-\theta_1)+0.5152=0\end{aligned} \quad (27)$$

Assuming different values for $(\theta_2-\theta_1)$ in equations 26 and 27, the corresponding values of θ_1 and θ_2 can be deduced, and from which another value of $(\theta_2-\theta_1)$ can be obtained. The difference between the assumed and deduced values of $(\theta_2-\theta_1)$ are tabulated against the assumed $(\theta_2-\theta_1)$. The required value is that which makes the corresponding difference equal to zero. In this example, θ_1 and θ_2 are found to be equal to 129.1 and 268.6 degrees, i.e., −91.4 degrees, respectively.

References

1. MATHEMATICAL ANALYSIS OF NONLINEAR CIRCUITS, **A. Boyajian.** *General Electric Review*, Schenectady, N. Y., vol. 34, 1931, pp. 531–37, 745–51.

2. CONTRIBUTION TO THE NONLINEAR CIRCUITS WITH LARGE APPLIED VOLTAGE, **W. L. Barrow.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 22, 1934, pp. 964–80.

3. OSCILLATORY CIRCUITS CONTAINING IRON-CORED INDUCTANCES, **Abd El-Samie Mostafa, M. El-Shishini.** *AIEE Transactions*, vol. 72, pt. I, July 1953, pp. 274–83.

4. RESONANCE THEORY OF SERIES NONLINEAR CONTROL CIRCUITS, **E. G. Keller.** *Journal*, Franklin Institute, Philadelphia, Pa., vol. 225, May 1938, pp. 561–76.

5. RESONANT NONLINEAR CONTROL CIRCUITS, **William T. Thomson.** *AIEE Transactions (Electrical Engineering)*, vol. 57, Aug. 1938, pp. 469–76.

6. NONLINEAR VIBRATIONS IN MECHANICAL AND ELECTRICAL SYSTEMS (book), **J. J. Stoker.** Waverly Press, Baltimore, Md., 1950, pp. 94–96.

Arc-Resistant Molding Materials and Finishes

R. F. STERLING
NONMEMBER AIEE

Synopsis: A means has been found to increase the arc resistance of certain organic materials which otherwise are inferior in this property. By incorporating any of a number of inorganic submicron refractory fillers, the American Society for Testing Materials (ASTM) arc resistance of phenolic and substituted phenolic molding compounds and a number of non-arc-resistant finishes can be greatly increased. Both the percentage and the particle size of the filler added are important factors. Data are presented describing the effect on arc resistance of these variables. A proposed mechanism presenting a possible explanation for the effects described in this paper is presented.

A COMMON TYPE of failure of organic electrical insulation is due to a conducting path forming on or just below the surface as a result of arcing between conductors affixed to the insulating surface. Resistance to this type of failure is called arc resistance and is defined in America by ASTM Standard *D495-48T*.[1]

Molding compounds based on phenolic resins have been used for molded insulation for many years. When compounded with various fillers and lubricants they are relatively low in cost and, when molded into a finished piece, possess many good electrical and mechanical properties. However, they have very poor arc resistance as measured by the ASTM test mentioned earlier, varying from 2 to 10 seconds with organic fillers up to 60 seconds with a glass filler.[2] This has seriously limited their use in electrical insulation, particularly in places where a flashover due to high-voltage surges or surface contamination may cause a conducting track and consequent failure of the molded piece.

Nature of Track Formation in Organic Materials

The nature of the conducting path formed in phenolic molding compounds as a result of failure under arcing indicates that this failure is due to elemental or low-molecular-weight electrically conducting carbon compounds being formed as a result of thermal or electrical degrada-

---

Paper 55-755, recommended by the AIEE Dielectrics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 3–7, 1955. Manuscript submitted August 8, 1955; made available for printing September 24, 1956.

R. F. STERLING is with Westinghouse Electric Corporation, East Pittsburgh, Pa.

The author wishes to express his appreciation to Dr. L. E. Edelman for his advice and efforts, particularly with regard to the work on finishes, and to Dr. G. R. Sprengling for consultation during the performance of the work described in this paper.

Table I. Effect of Particle Size of Silica Additive on Arc Resistance of Moldings Made from Wood-Flour-Filled Phenolic Resin

Trade Name	Particle Diameter, Microns	Per Cent Added	ASTM Arc Resistance, Seconds
PPG Hi-Sil	0.02	10	103.4
JM Snow Floss	1–2	10	51
JM Super Floss	2–4	10	33
Novacite 2500	5	10	17.6
Blank, no additive			7

Table III. Effect of Low Percentages of Several Submicron Fillers on Arc Resistance of a Commercial General-Purpose Molding Compound*

Ingredient	Per Cent in Molding Compound	Arc Resistance, Seconds
Hi-Sil (trade name for Pittsburgh Plate Glass Company silicon dioxide. Mean particle diameter, 0.02 micron)	8	124
	7	124
	6	111
	5	100
	4	85
	3	93
	2	76
	1	57
	0	44
ASP-100 (trade name for aluminum silicate pigment marketed by Edgar Brothers Company, Metuchen, N. J. Mean particle diameter, 0.5 micron)	5	98
	4	108
	3	79
	2	73
	1	69
	0	44

*Molding compound composition: 62.30% novolak resin, hexa, catalyst, and lubricants; 35.98% dried wood flour; 1.72% nigrosine dye.

Table IV. Effect of Particle Size of Silica Added to Styrenated Epoxy Resin Ester Enamel

Material	Particle Diameter, Microns	Per Cent Added	ASTM Arc Resistance, Seconds
PPG Hi-Sil	0.02	10	182
Columbia Southern (85–1207)	0.02	10	183
Columbia Southern (109–354)	0.03	10	183
Columbia Southern (85–1073)	0.08	10	183
Novacite 2500	5.0	10	76
Novacite 625	20.0	10	59
Blank, no additive			62

tion. The very high carbon content, about 80 per cent (%) by weight, and extent of carbon-to-carbon bonding render phenolics particularly susceptible to arc failures. This relationship between a phenyl ring structure and low arc resistance was noted by Warren[3] who emphasized the superior arc resistance of alkyl groups to phenyl groups when substituted on silicone polymer chains.

Olyphant,[4,5] using high-speed photographic methods, studied the tracking process ("tracking" in this case refers to a conducting path due to an external arc) and the effect on this process of various testing conditions. He examined the mechanisms of both track initiation and track propagation in a number of materials, including mineral- and wood-flour-filled phenolics, alpha-cellulose- and asbestos-filled melamine, alpha-cellulose-filled urea, glass-bonded mica, and a series of compressed filler materials.

From his work he concluded that if a carbonized area can be readily formed in the piece under test in intimate contact with the electrode, an artificial extension of the electrode into the material can take place and track initiation is facilitated. Since tracking, once started, is usually rapid, factors which influence the initiation of the track are the most critical in determining arc resistance. He further concluded that the high (130-second) arc-resistance values obtained for amino resins is due to the formation of cavities next to the electrodes which insulate the carbonizing areas and prevent the direct electrode extension. In fact, by the simple means of raising the electrodes as little as 0.01 inch from the surface of the test piece, he was able to obtain arc-resistance values for phenolic molding materials of 120 seconds or more, while under the same conditions of raised electrodes melamine molding materials still failed between 130 and 140 seconds.

Experimental Procedure and Results

When the problem of the development of new arc-resistant molding materials was undertaken at Westinghouse Research Laboratories beginning in 1951, it was felt that if one could interpose atoms which would be nonelectrically conducting in the carbon-to-carbon chain, one would expect chain conductivity to be lowered, and arc resistance to be greatly increased. Evidence for this is found in the high arc resistance of amino resins, where nitrogen is interposed between carbon atoms in the polymer chain. It was felt that a worth-while approach would be to consider the introduction of inorganic atoms or molecules into the carbon polymeric chain.

Following this thought, a literature search disclosed that previous investigators, especially Hendricks,[6] had reacted aniline with montmorillonite, a complex inorganic molecule, to form a compound in efforts to obtain a polymeric soil stabilizer. If a copolymer could be formed of aniline montmorillonite, phenol, and formaldehyde which was moldable under heat and pressure, it might have high arc resistance, due to the relatively bulky inorganic montmorillonite molecule preventing a continuous carbon path from forming after arcing.

EXPERIMENTAL PROCEDURE

Following Hendricks procedure, aniline montmorillonite was made and identified. This, when blended with modified phenolic resins previously made, showed good molding properties and arc resistances of over 120 seconds.

The question then arose as to the possibility that a material with characteristics similar to montmorillonite in refractory, nonelectrically conducting, and especially very fine particle dimensions, might show the blocking or interposition effect previously mentioned without actually forming a primary valence bond with carbon on a phenolic ring. Boron nitride, a very finely divided high-melting electrical insulator which was available commercially, was among the first such materials tried. This showed a great increase in arc resist-

Table II. Effect of Particle Size of Aluminum Silicate Fillers on Arc Resistance of Moldings Made From High-Temperature Mineral-Filled Phenolic Molding Compound

Trade Name	Chemical Composition	Particle Diameter, Microns	Per Cent Added	ASTM Arc Resistance, Seconds
ASP-100*	X Al_2O_3. YSiO_2	0.5	10	176
ASP-500*	X Al_2O_3. YSiO_2	0.8	10	178
ASP-900*	X Al_2O_3. YSiO_2	1.5	10	161
ASP-400*	X Al_2O_3. YSiO_2	4.2	10	144

*These are all the same chemical composition, having typical analysis as follows: silicates as SiO_2, 44.20%; alumina as Al_2O_3, 40.14%; ignition loss at 950 C, 13.68%; trace oxides (TiO_4, Na_2O, M_1O, etc.), 0.94%; moisture at 100 C, 0.32%.

Table V. Effect of 10% Hi-Sil (0.02-Micron-Diameter Silica) on Arc Resistance of Several Insulating Varnishes

Material	ASTM Arc Resistance, Seconds: Blank, No Additive	ASTM Arc Resistance, Seconds: 10% Hi-Sil
Phenolic varnish	14	86
Styrenated epoxy resin ester + 20% phenolic	61	137
Styrenated alkyd	136	162

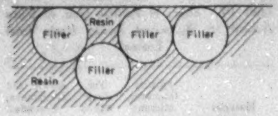


Fig. 1. Hypothetical magnified cross section of molding using conventional fillers

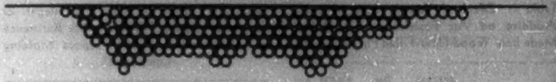


Fig. 2. Hypothetical magnified cross section of molding using submicron filler

ance, even better than the aniline montmorillonite modified materials had shown. The work was then extended, first to other resins, then to other submicron-diameter refractory fillers, until the fact was established that quite a number of these materials would improve ASTM arc resistance in phenolic molding compounds. The work was later extended to cover coatings and finishes.

The procedure followed in compiling the data shown in Tables I–III report was as given in the following.

The molding resin was first compounded. This was done in either of two ways. The first way was simply to dry-mix the submicron fillers into a commercial wood-flour or asbestos-filled molding material and then to compound these fillers into the molding material by milling on steam-heated differential rolls at temperatures of 100–115 C (degrees centigrade). The second way was to start with either a commercial or a laboratory-prepared novolak and add the required amount of hexamethylenetetramine, lubricants, and auxiliary fillers to give the moldability and mechanical properties desired before milling. It was found very early in the program that this milling was necessary to insure uniform distribution on a micro scale of the filler into the molding compound. The resultant blanket of molding material as removed from the rolls was then broken up and ground in a Wiley hammer mill to a fineness passing through a 16-mesh screen. The resultant coarse powder was then placed in airtight glass jars and stored until molding. The molding itself was largely performed on a hand-operated Carver laboratory press in a 2-inch disk mold, the disks being 2 inches in diameter and about 1/8 inch thick. More promising formulations were then molded into pieces of operating equipment such as circuit breakers and meter bases for further testing.

ASTM arc-resistance tests were carried out on the various pieces on the testing equipment at Westinghouse Research Laboratories. The more promising specimens were checked by other operators on the ASTM equipment at the Westinghouse Engineering Laboratories.

Experimental Results

It can be observed from Tables I and II that the improvement in ASTM arc resistance of otherwise non-arc-resistant phenolic molding compounds is roughly inversely proportional to the particle diameter, with slight improvement detectable at 5-micron particle diameter and increasing arc resistance as the ultimate particle diameter of the filler is reduced to 0.5 micron or less. Table III shows that a beneficial effect on arc resistance is detectable when the fillers are added in quantities as low as 1% by weight. Composition of the finely divided filler is apparently not critical if requirements of thermal stability and electrical insulation are met. Oxides of silicon, aluminum, zirconium, titanium, and beryllium are available in submicron particle diameters, and have all demonstrated this effect on ASTM arc resistance, as well as the aniline montmorillonite and boron nitride mentioned earlier. For a given particle size and percentage of a given filler, the improvement in arc resistance is affected by the other fillers present. For example, 10% of aluminum silicate of 0.5-micron mean particle diameter increases the arc resistance of wood-flour-filled phenolic resins to 120–130 seconds and of asbestos-filled phenolic resins to over 180 seconds.

Because of a lack of availability of a range of particle sizes of many of the refractory materials, the tables illustrate the effect on arc resistance of silica and aluminum silicate, which are available in a well-classified range of particle sizes.

The effect of some of these fillers on the arc resistance of various finishes is shown in Tables IV and V.

Discussion of Results

The following explanation seems probable for the results described in this paper. The ASTM test for arc resistance utilizes a high-voltage low-current arc, with the electrodes so placed and shaped that the arc drawn between them is bowed upward and failure in phenolic-molded products always occurs by initiation of a conducting track at one of the electrodes, followed by rapid extension of what is then a point electrode across the surface of the molded piece. Olyphant's study had shown that in phenolics track growth occurred only on

Fig. 3. Conducting track immediately after initiation at electrode on left

Fig. 4. Conducting track completed

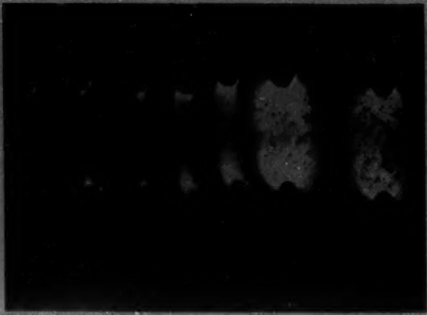


Fig. 5. Progressive formation of refractory mat on phenolic molding after 15, 30, 60, 90, 120, and 180 seconds, and failure

Fig. 6. Same molding as Fig. 5, lighted to illustrate interference patterns formed probably due to erosion of organic molding by arcs

the half-cycle when the advancing point forming the conducting track was negative with respect to the stationary electrode, the arc remaining fixed in length on the half-cycles when the advancing point forming the track was positive. If, however, track initiation can be prevented, track growth cannot begin. To prevent track initiation in phenolic resins as described in this paper, not only must the refractory filler be of sufficiently fine particle diameter, but it must also be uniformly distributed in the molding compound. It has therefore been found necessary to incorporate the filler by milling on hot differential rolls to obtain the desired high arc resistance. Dry blending of these fillers into the molding resin has proved inadequate to obtain uniformly high arc resistance.

If the rate of degradation of the phenolic polymer at the electrode tip can be reduced to the point where all the products of degradation of the carbon-rich phenolic are oxidized and volatilize off as gaseous carbon monoxide or carbon dioxide, track initiation can be prevented. This point can be illustrated by Figs. 1 and 2. Fig. 1 is a hypothetical cross section of a molded phenolic piece using conventional filler. Fig. 2 is the same cross section using an equal weight of submicron filler. The fillers in both cases are assumed to be roughly spherical in shape.

Although, assuming that the shapes of large and small particles of filler are the same, the resin-to-filler ratio is the same in the two illustrations, it will be seen that at any time there is a much larger and thicker volume of pure resin available to break down thermally and form a conducting carbonaceous residue in the case illustrated by Fig. 1 than that illustrated by Fig. 2. If the pieces illustrated are subjected to the ASTM arc-resistance test, the arc, particularly near the electrode tip, rapidly decomposes the resin present on the surface of the molding. In the case shown in Fig. 1, the resin breaks down thermally to low-molecular-weight carbonaceous residues which, because of insufficient oxygen present, do not vaporize and therefore remain in place and very rapidly form a conducting path. This is shown in Fig. 3, where the arc has carbonized a portion of the resin immediately adjacent to the electrode. This conducting portion then forms an extension of the electrode and the track will, in a few more cycles, extend to the other electrode, giving the characteristic thin wire-like track shown in Fig. 4. The spots of greater decomposition on the track are those described by Olyphant[4,5] as due to the half-cycle in which the track is positive with respect to the other electrode and not growing, track growth only occurring on the half-cycles when the track is negative to the other electrode. In the case shown in Fig. 2, however, the resin immediately adjacent to the electrode tips will again break down thermally to low-molecular-weight carbonaceous residues when the arc is applied. However, because of the presence of these residues in relatively thin films between particles, oxidation occurs and the residue is vaporized. A second beneficial effect is observed as this process continues under repeated arcs. With the vaporization of the inorganic material present, the inorganic material appears to remain in place, forming a porous insulating mat of refractory material. This may be seen in Figs. 5 and 6 which show a conventional asbestos-fiber- and wood-flour-filled phenolic to which 10% by weight of boron nitride was added. Fig. 5, with the photograph taken normal to the surface, shows the mat formation, beginning at the left, after 15, 30, 60, 90, 120, and 180 seconds, and failure, which occurred at 185 seconds. This failure occurred by a track forming around the mat at the edge, at the extreme right side of Fig. 5.

Fig. 6 shows the first six tracks of the same test piece. It is an attempt to indicate the thickness of the insulating mat remaining after the organic portion of the molding is vaporized. An interference pattern, characteristic of thin films, appears first near the electrodes and then gradually extends toward the other electrode. As shown in Fig. 6, an interference pattern can be seen between the resin subsurface and the refractory mat surface after 15, 30, and 60 seconds of the ASTM test, but after 90 seconds the film is so thick that no interference can be discerned except at one edge.

Conclusions

A means has been found to increase the arc resistance of certain organic materials which otherwise are inferior in this property. By incorporating any of a number of inorganic submicron refractory fillers, the ASTM arc resistance of phenolic and substituted phenolic molding compounds and a number of non-arc-resistant finishes can be greatly increased. A postulated mechanism explaining the observed phenomenon is proposed. The major conclusions to be drawn from this work are as follows:

1. The addition of inorganic refractory fillers of sufficiently fine particle diameter is a means which can be used to increase the ASTM arc resistance of phenolic and substituted phenolic resins, as well as certain otherwise non-arc-resistant finishes. These fillers, which include oxides of silicon,

aluminum, beryllium, zirconium, and titanium, combinations of these in the form of silicates, titanates, and zirconates, as well as boron nitride and aniline montmorillonite, are, for the most part, quite inexpensive, and when added in amounts up to 10% by weight, do not seriously affect the mechanical properties of the molded piece.

2. The improvement in the ASTM arc resistance of phenolic and substituted phenolic molding compounds, as well as certain otherwise non-arc-resistant finishes, is an inverse function of particle diameter of these refractory inorganic fillers. At particle diameters above approximately 5 microns, the effect described essentially disappears.

3. The beneficial result of these fillers is a function of percentage content in the molded article or finish. Beneficial results were observed at as low as 1%-by-weight filler; however, 10% addition provided considerably greater arc resistance.

4. In use of this means of improving ASTM arc resistance, the importance of uniform dispersion of the filler in the molding compound or enamel is critical. Differential roll milling is required to obtain the effects described in this paper.

5. Molding materials and finishes possessing high arc resistance intrinsically, such as the amino resins, are not markedly improved in this quality by addition of the fillers described in this paper.

6. The data and conclusions presented in this paper are all based on results obtained on equipment testing to ASTM Standard *D495-48T*, with currents at 10 milliamperes, and effects of a range of higher currents such as are involved in power apparatus have not been investigated. No equipment has been standardized or has achieved wide usage for testing arc resistance at higher power levels.

References

1. *Standard D495-48T*. American Society for Testing Materials, Philadelphia, Pa.

2. MODERN PLASTICS ENCYCLOPEDIA. Plastics Properties Chart, Plastics Catalogue Corporation, New York, N. Y., 1953.

3. INSULATION, H. Warren. *Journal*, Institution of Electrical Engineers, London, England, vol. 87, no. 528, 1940, p. 588.

4. ARC RESISTANCE: I–TRACKING PROCESSES IN THERMOSETTING INSULATING MATERIALS, M. Olyphant. *Bulletin*, American Society for Testing Materials, Philadelphia, Pa., Apr. 1952, pp. 60–67.

5. PART II. *Ibid.*, Oct. 1952, pp. 31–38.

6. BASE EXCHANGE OF THE CLAY MINERAL MONTMORILLONITE FOR ORGANIC CATIONS, S. B. Hendricks. *Journal of Physical Chemistry*, Washington, D. C., vol. 45, 1941, pp. 65–81.

A Transistorized Time-Division Multiplex Telegraph Set

F. D. BIGGAM
ASSOCIATE MEMBER AIEE

Synopsis: Increased usage of radio-teletypewriter communication has widened the demand for facilities for multiplexing teletypewriter channels on a single radio channel, for economy of frequency spectrum and equipment. Many of the potential applications of multiplex require smaller and lighter equipment than heretofore available, with unusually high reliability. This paper describes a developmental time-division multiplex equipment which, through the use of transistors, has only 20 per cent of the volume, 20 per cent of the weight, and 5 per cent of the power consumption of the equivalent electron-tube set now in service. The set is packaged for maximum operating convenience and serviceability in the field.

THE transistorized multiplex is a completely transistorized 4-channel time-division multiplex telegraph terminal set. It is functionally similar to, and compatible with the *AN/FGC-5* electronic multiplex,[1] which has been in general use since 1950 on long-haul high-frequency radio circuits. In all active electronic functions such as switching and amplification, as well as in the power supply, only solid-state devices are employed. There are no vacuum tubes in the set.

The equipment is designed as an integral part of a time-division multiplex system wherein two, three, or four start-stop telegraph channels are combined for transmission over a common communication circuit. It occupies approximately one fourth of the physical volume of the vacuum-tube set, as is apparent in Fig. 1. The complete send-receive terminal is housed in one 36-inch-high cabinet and weights only 275 pounds, as compared to 1,450 pounds for the electron-tube equivalent. It consumes only 1/20 as much electric power in order to perform identical functions. The weight, size, and power savings are made possible by the application of transistors and solid-state devices. The set contains over 550 transistors and over 700 germanium diodes.

History

The use of time-division multiplex equipment in commercial telegraph service dates back as early as the turn of the century. The original multiplexes were motor driven and mechanically operated. Electromechanical systems were subsequently adapted for use on radio-telegraph circuits during World War II, where the multiplexing of channels served both to increase circuit capacity and to make more efficient use of the frequency spectrum. With the development of frequency-shift keying for radio transmitters, and of improved radio-receiving systems, interest in time-division multiplex was further stimulated. Since electromechanical designs were difficult to maintain, were hard to phase, and had a low order of frequency stability, a new approach was dictated. The electronic multiplex in general use today was developed as a result of this need. Over 500 of these sets have been manufactured. They are presently in service throughout the world, capably performing the function for which they were designed.

More recently, because of further improvements which have been made in transmission techniques, and of even greater traffic loads, new design requirements for time-division multiplex facilities have been created. In some applications, an equipment of minimum size, weight, and power consumption is demanded. Among these are mobile and shipboard installations. Investigations were therefore made into means for better meeting these requirements, which were imposed by this potential demand. As a result the transistorized multiplex equipment described here was designed; four developmental models are presently undergoing field trials.

The Multiplex System

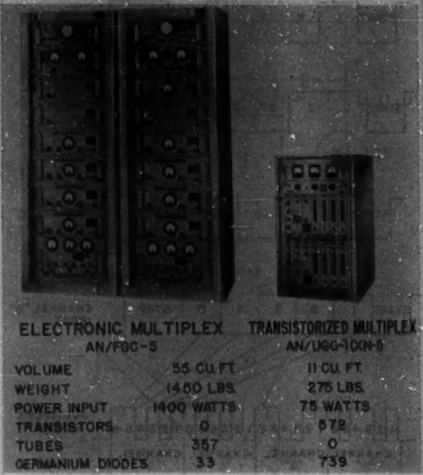
The transmitting multiplex responds to input telegraph signals of the standard 7.42-unit (Baudot) code. These signals may be supplied either from a continuous

---

Paper 56-987, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted June 8, 1956; made available for printing August 13, 1956.

F. D. BIGGAM is with the Teletype Corporation, Chicago, Ill.

The equipment described in this paper was developed through the combined efforts of several numbers of Teletype's technical staff: T. A. Hansen, project engineer, R. J. Reek, R. A. Slusser, and P. G. Wray, circuit design; H. G. Klassen and the author, packaging.



	ELECTRONIC MULTIPLEX AN/FGC-5	TRANSISTORIZED MULTIPLEX AN/UGC-1(XN-1)
VOLUME	55 CU. FT.	11 CU. FT.
WEIGHT	1450 LBS.	275 LBS.
POWER INPUT	1400 WATTS	75 WATTS
TRANSISTORS	0	572
TUBES	357	0
GERMANIUM DIODES	33	739

Fig. 1. Comparison data, telegraph terminal sets

transmitting device such as a transmitting distributor, or from a teletypewriter keyboard. Fig. 2 shows four typical input signals. Each input character consists of a start element followed by five code pulses and a stop element. The start and stop elements are always the same, "space" (no current) for the start pulse and "mark" (current) for the stop pulse, while each of the five code elements may be either mark or space, depending upon the character being transmitted. The input signals are fundamentally nonsynchronous. More accurately, successive pulses within any one character are transmitted and received synchronously, but the timing of successive characters is not synchronous, and may be completely random. The function of the start and stop pulses is to resynchronize, or rephase the receiving printer with the received signal at the beginning of each character. Other than to establish a reference point, the start and stop pulses contain no information.

The multiplex, on the other hand, is a fully synchronous system. The receiving terminal is maintained in step with the transmitting terminal, in both frequency and phase, by a frequency standard which is independent of the signal being transmitted. There is therefore no need to transmit synchronizing information at the start of each character. Consequently, in forming the multiplex signal only the five code elements are used, the start and stop pulses having been dropped at the transmitting end. A sixth pulse is added by the multiplex, the function of which will be described. At the receiving terminal, the signal is reconstituted by the automatic insertion of the start and stop pulses.

Since start-stop inputs to the multiplex may be of slightly varying transmission rates and have no direct phase relationship between channels, it is necessary that multiplex transmission be slightly faster than the incoming start-stop signals. This results in a constantly sliding or varying phase relationship between the start-stop and multiplex signals, with the multiplex signal continuously gaining on the start-stop signal. Occasionally the multiplex signal will get sufficiently ahead of the start-stop signals so that it must pause to allow the start-stop signal to catch up. This is accomplished by the automatic insertion of multiplex "blank," which is an all-spacing group. Since the blank combination is considered a true character in the start-stop code, it is necessary for the multiplex receiver to distinguish between code blanks which occur when the channel is idle, or because of this speed difference, and blanks which should be transmitted over the channel and delivered to the distant receiving circuit, because they were among the start-stop signals received from the originating transmitters. To accomplish this, a sixth element has been added to the multiplex signal code. This element is made marking for each transmission of character combination, and spacing for each transmission of an idle signal, or time differential blank.

Since four characters must be transmitted over a single circuit during the time ordinarily required by one, they must be compressed with respect to time so that each channel is transmitted during approximately one fourth of the time necessary for its initial generation. Fig. 2 shows the multiplex signal formed by the four start-stop inputs. Each channel is composed of fine code elements followed by the sixth pulse previously discussed. The multiplex signal shown is further modified before transmission by the inversion of certain pulses. These pulses are sent out as space when intended to be mark, and mark when space, in order to provide mark-space transitions during transmission of an idle signal. These transitions are necessary for synchronizing purposes. At the receiving terminal they are reinverted to their proper polarity.

A simplified block diagram of a transmitting and receiving multiplex system is shown in Fig. 3. At the left side, telegraph transmitters deliver start-stop signals to the individual transmitting group telegraph code converters. It is the function of the code converters to assemble the input information, which is in serial form, and present it to the multiplexer in parallel binary form. The multiplexer sweeps over the outputs of the code converters, channel by channel, picking up a complete character in turn from each converter at the multiplex frequency rate. The multiplex operating speed is controlled by the frequency standard, which has as its heart a transistorized crystal oscillator circuit. Impulses generated by the crystal oscillator and the various dividing circuits are used to step the multiplexer at the proper rate. By varying the pulse division ratio, the equipment may be switched to operate with inputs of either two, three, or four channels, or to accommodate telegraph speeds of 60, 75, or 100 words per minute, depending upon traffic requirements and radio-frequency propagation conditions. Multiplex outputs may be either keyed tone or 60-milliampere d-c signals. These signals are applied to the external transmitting equipment.

At the receiver, the demultiplexer performs functions complementary to those of the multiplexer. It separates the information and routes it to the proper code converter. The converters examine the input signal and, provided a character transmission has been recognized, begin the process whereby the start-stop signal is reconstituted and regenerated. The sixth

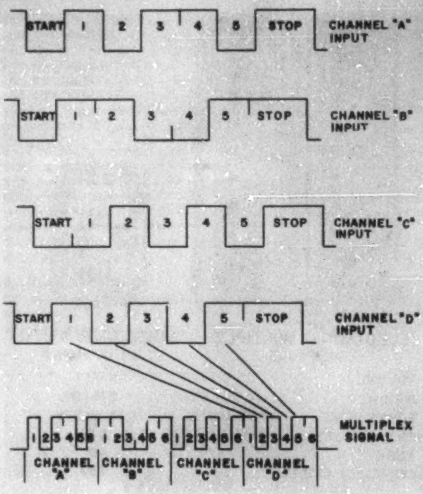


Fig. 2. Composition of the multiplex signal: Four start-stop signals are compressed with respect to time and transmitted channel by channel

pulse, which is used only to recognize the presence of a bona fide blank transmission, is dropped. The start and stop elements are added to the five code pulses and the signal is read out at telegraph speed.

The frequency standard at the receiving station performs a function identical to that of the transmitting group, except that its effective drive must be modified in the receiving group in order to maintain synchronism with the transmitting group. This is accomplished by a unique synchronizing circuit which operates on a digital basis. By noting the time location of space-to-mark transitions of the incoming multiplex signal, the exact phase relationship between the incoming signal and the demultiplexer drive is established. Synchronizer action is accomplished by adding or subtracting drive pulses in discrete steps until a proper time-phase relationship is achieved. This system of synchronizing permits the use of a single crystal oscillator to drive both the transmitting and receiving groups in a given location, since the frequency of the crystal oscillator remains fixed. The correction rate may be noted on the front panel by the flashing of a neon lamp. The frequency of these corrections is limited to a maximum value, which may be varied from 6.25 pulses per second, to one pulse per 20 seconds, by means of a synchronizing rate-control switch. The action of the synchronizer may therefore be slowed during poor signal periods. Since the synchronizer responds to changes in phase relationship between mark-space signal transitions of the incoming multiplex signal, it will attempt to establish as in-phase relationship, not only with normal signals, but with all extraneous bursts of noise and distortion which may be received. By limiting the maximum number of correction pulses generated by the synchronizer per unit time, it is possible to average out effectively these extraneous pulses, correcting only for frequency drift between the transmitting and receiving crystal oscillators.

In addition to the establishment of a time-phase relationship between the incoming signal and the demultiplexer drive, the receiver must be placed in proper channel register with the transmitter. That is, the proper information must be routed to the proper channel by the demultiplexer. With *A* channel transmission disabled, operation of a phasing button causes a train of drive pulses to be generated which add to the demultiplexer drive in order to advance its phase. The phasing pulses are automatically stopped when proper channels come into register.

Circuit Description

Transistors have characteristics which make them particularly suitable for use in equipments of this type. As switches, their very high resistance when "off" and practically zero resistance when "on," make them extremely useful, and as storage elements they store information at very low power levels.

Point-contact transistors are used in the equipment along with the more common junction type. Characteristics of point-contact transistors, which enable the realization of emitter to collector current gains greater than unity, greatly simplify the design of certain of the circuits.

There are several basic circuits which are employed to accomplish the necessary switching and logic functions. Among them are the ring distributor, storage, start-stop oscillator circuits, and the power supply. Since these circuits are typical of those employed, they will be described in the following paragraphs.

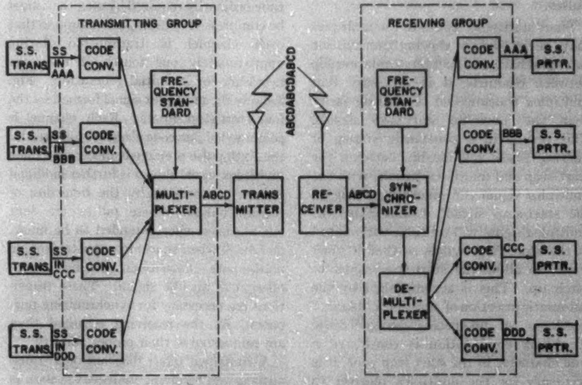


Fig. 3. Simplified block diagram, multiplex terminal set

RING DISTRIBUTOR

The ring distributor may be thought of as the basic circuit of the transistor multiplex. In addition to its basic distribution function, it is employed in the equipment as a frequency divider. There are 21 separate ring distributors in the equipment. A 3-element ring is shown in Fig. 4. While only three elements are shown, the ring may be connected in any configuration ranging from two to a very large number. Rings of this type have been reliably operated with as many as 26 stages, and indications are that an even greater number of stages could be connected if desired. However, the number of stages per ring varies from three to seven in the transistor multiplex.

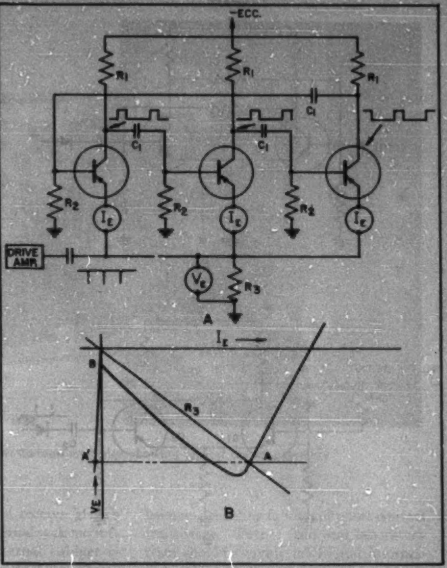


Fig. 4. Ring distributor

A—3-element ring distributor

B—Emitter input negative resistance characteristic of ring distributor element

The point-contact transistor, which is a highly nonlinear device for large signal operation and has an emitter-collector current gain greater than unity, is utilized in the basic ring circuit to accomplish the sequential switching function. With power applied to the circuit of Fig. 4(A), the distributor element with the highest combination of reverse collector-base current (I_{co}) and emitter current amplification (alpha) conducts into saturation. Current flowing through the common emitter resistor *R3* develops a voltage which when applied to the emitters of the off elements, biases them sufficiently in the reverse direction to prevent their conduction. Fig. 4(B) shows the input negative resistance characteristic curve of emitter voltage and emitter current for a typical ring element. When conducting, the element is stable at point *A*. Because the emitter resistor *R3* is common to both on and off elements, the off elements are held at point corresponding to a very slightly negative emitter current. The off ring elements are thus maintained in that condition by a bias voltage equal to *BA'*.

The distributor is stepped from one conducting element to the next by the following means: A negative drive pulse is applied to the common emitter connection which causes the conducting element to switch to the nonconducting state. When the collector voltage of the on element switches from slightly negative to highly negative, a differentiated negative wave form is passed through *C1*, and impressed on the base resistor of the following stage. When the stepping or triggering pulse is removed, the primed stage conducts into saturation.

This circuit, because of its basic simplicity, has made the application of point-contact transistors very practical. Only slightly over three components per stage, besides the transistors, are required.

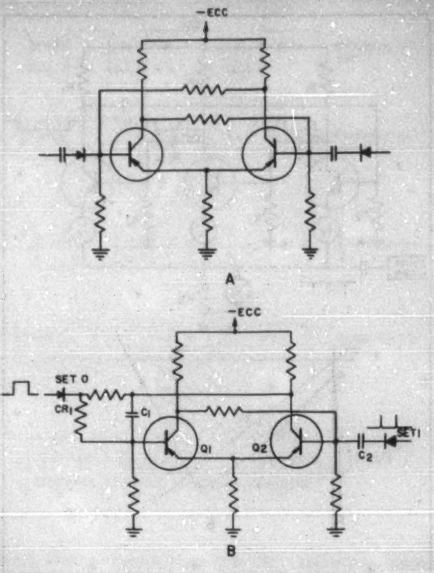
Because the circuit is arranged to be basically monostable, and therefore not dependent upon the interception of the emitter load line (*R3*) more than once with the emitter input characteristic curve of Fig. 4B, very satisfactory operating margins are achieved without adjustment of the circuit constants. No stabilization circuits are required. This circuit was designed to operate with transistors which have emitter-to-collector current gains as low as 1.8, and which allow reverse current flow (I_{co}) at 20 volts and 60 degrees centigrade to be as high as 3 milliamperes. The −20-volt operating voltage permits a 10-volt signal excursion between off and on, and forms the standard to which the voltage swings of all other circuits are matched.

STORAGE

Binary storage of information is accomplished by the use of Eccles-Jordan *p-n-p*-type junction transistor flip-flops as shown in Fig. 5(A). All storage, and many of the switching and timing circuits, are of this type. Triggering is accomplished by a positive going pulse applied to the base of the conducting element driving it to its off condition. The circuit is designed so that output levels will be compatible with that of the ring distributor circuit previously discussed. Voltage swings from −6 volts in the on to −16 volts in the off condition are therefore obtained.

Junction transistors used are generally low-frequency response units having alpha cutoff frequencies of 1 megacycle. Switching requirements were not sufficiently severe to warrant the use of higher frequency units. Grounded emitter current gains of as low as 30 may be tolerated in the circuits. The maximum reverse current at room temperature was specified as 15 microamperes.

A unique application of the basic storage circuit is shown in Fig. 5(B). At certain periods during operation of the equipment it becomes necessary for this circuit to act either as a bistable flip-flop, or as a monostable univibrator. Ordinarily the triggering, or set *1*, drive arrives prior to the reset, or set *O* pulse, in which case the circuit functions as a normal binary. However, at other times the set *O* pulse actually arrives prior to the set *1* pulse. It is necessary during these periods for the reset pulse to prime the stage in order that, upon the receipt of the triggering pulse, an output is obtained, and, after a timed interval, the circuit is automatically returned to the original condition. The desired function is controlled by the potential applied to the circuit through diode *CR1*. When the



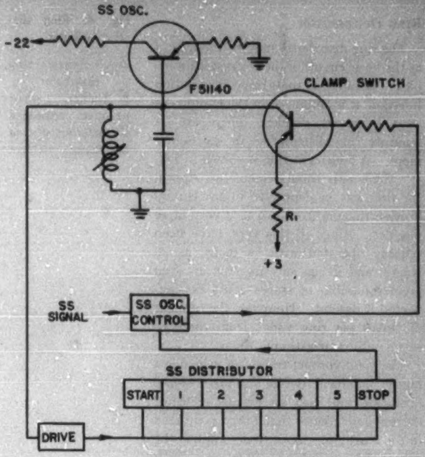


Fig. 5 (left). Binary storage

A—Binary storage element
B—Modified binary, bistable-monostable circuit

Fig. 6 (above). Start-stop oscillator and control

potential applied is most negative, *CR1* is biased in the reverse direction and is prevented from affecting the operation of the circuit. When a drive pulse is applied to the base of *Q2*, assuming *Q2* is conducting, the circuit will trigger, with *Q2* becoming nonconducting and *Q1* conducting. The flip-flop therefore operates in the conventional manner.

When the potential applied to *CR1* swings from its most negative to its least negative value however, *CR1* becomes forward biased, effectively disabling the cross-coupling circuit. Capacitor *C1* then serves as the only cross coupling between the collector of *Q2* and the base of *Q1*, causing the circuit to become a univibrator. Upon the application of a positive trigger pulse to the base of *Q2*, the circuit triggers. After a timed interval, determined by the charging time of *C1*, it is restored to its original condition.

START-STOP OSCILLATOR

The basic timing means, whereby the signal is converted from sequential start-stop to parallel 5-wire form at the transmitter, and conversely at the receiver, is the start-stop oscillator. Each character of the start-stop signal being received by the code converters at the transmitting terminal triggers the start-stop distributor, which then samples the incoming signal. Once started, the start-stop oscillator oscillates through six cycles and comes to rest of its own accord. Fig. 6 shows the oscillator and clamp, together with associated circuitry, in simple block diagram form. The basic oscillator is of the negative resistance type with an inductance/-capacitance (I/C) tank connected in the base circuit of a point-contact transistor. The frequency of oscillation is such that one period is equal in time duration to that of one code pulse in the start-stop signal. For example, at 60 winds per minute the period of oscillation is 22 milliseconds, which corresponds to a frequency of 45.45 cycles per second. The output is applied to the distributor drive circuits in order to step the start-stop distributor. Beginning with the start element, the distributor will cycle through each of its seven elements at the oscillator frequency rate. When next the stop element fires, a signal will be fed to the control circuits which will in turn cause the clamp switch to become effectively a short circuit across the oscillator tank. Oscillations will then be clamped until the next start pulse is received by the control circuits. This oscillator is inherently stable within ±1 per cent over a 0–60-degree-centigrade temperature range.

POWER SUPPLY

The power unit is an unregulated multiple-voltage supply, consisting of a power transformer, germanium full-wave rectifiers, and I/C filtering. The principal operating voltage is −22 volts. This voltage supplies the ring distributors, storage elements, and nearly all other circuits. In addition, a separate variable −22-volt supply has been included in the system as a test facility for voltage margin measurement. It is possible to vary the −22-volt supply in each drawer, without affecting the rest of the equipment, in order to detect marginal circuits.

The equipment is designed for 115-volt 60-cycle operation. However, all circuits will operate with input variations of from 90 to 125 volts, 50 to 60 cycles.

Packaging

Extreme miniaturization was not the primary objective in packaging the equipment, although the transistorized multiplex occupies only approximately 25 per cent of the total volume of the vacuum-tube send-receive set. Reliability and serviceability were considered of primary importance, but consistent with these, miniaturization was to be desired. To this end, half-watt 5-per-cent carbon resistors and metalized paper capacitors with 200-volt d-c ratings were adopted as standard. Transistor sockets were provided for replacement of the transistors as a conservative approach to the mounting problem, even though a high degree of transistor reliability was anticipated. A combination of proved mounting techniques and etched circuit wiring was used. Conventional terminals were employed for securing circuit components, which facilitate replacement of parts in the field. All circuit elements are mounted on etched circuit cards and are spaced conservatively, with adjacent components not mounted closer than 0.3 inch center to center. A typical drawer assembly is shown in Fig. 7.

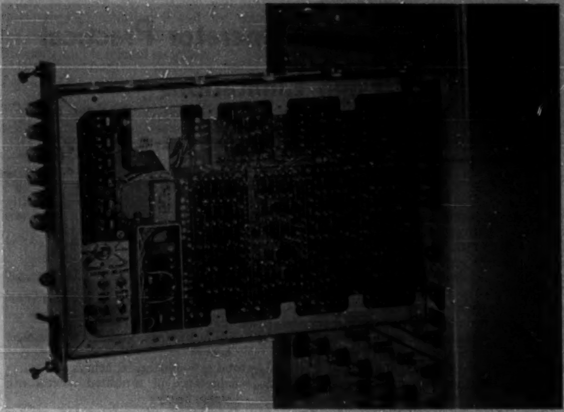


Fig. 7. Transistorized multiplex showing typical drawer assembly

Several circuit cards are mounted on a larger etched wiring panel, and connected to it by wire straps and conventional soldered terminals in order to facilitate removal and replacement of the cards as modular units, without introducing a large number of friction contact connections. The wiring panel thus serves as a chassis and local cable. It is mounted vertically in a drawer assembly, which may be withdrawn from the cabinet as shown in Fig. 7. All components are visible and readily accessible when the individual drawers are withdrawn.

There are 14 drawers in the system, in addition to the power supply. They are comprised of seven basic types, which slide into vertical slots in the send and receive group enclosures. A novel collapsible cable assembly has been provided which allows the drawers to be withdrawn without interrupting traffic while the equipment is operating. Convenient equipment accessibility is thereby achieved without the necessity of connecting adapter cables and plugs.

This method of packaging for convenience of maintenance was adopted as a conservative design approach. In view of the low voltages and operating temperatures throughout the equipment, need for servicing and replacement of components is expected to be infrequent.

The cabinet is of aluminum construction throughout. It is designed to permit removal of the send and receive groups and power supply for external rack mounting if desired. The external cabinet is therefore basically no more than a protective cover for the equipment. All interconnections between local send and receive groups and between drawers are accomplished via cross-connecting panels located at the rear of the group enclosures, and by patch cords between the group enclosures and power supply.

Operation

The transistor multiplex, while still in the developmental stage, has shown promise of providing extremely high reliability in operating service. Transistor and component failures traceable to circuit operation have been almost nonexistent. Some breadboard circuitry has been operating for as long as 2 years without failure. Those few failures which have occurred have been in junction transistor circuits and were traceable either to increased reverse current, or loss in current gain. To date, there have been no operating failures of point-contact transistors. Future field and laboratory tests should provide additional information which will permit more accurate prediction of the ultimate reliability to be expected of the equipment.

Conclusions

Application of solid-state devices to the multiplexing function, enabling the reduction of volume and power requirements, should provide new applications for the transistor multiplex. Decreasing transistor and diode costs indicate that when it is produced in quantity, the transistor multiplex will be competitive in price with the vacuum-tube multiplex. What is most important, however, is that the expected long life of transistors should provide a degree of reliability so far unobtainable in electronic equipment.

Reference

1. An Electronic Time Division Multiplex Telegraph Set, T. A. Hansen, R. D. Slayton. *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 354–61.

Magnetic-Pulse-Generator Practical Design Limitations

M. F. THOMPSON
NONMEMBER AIEE

R. R. TRAUTWEIN
ASSOCIATE MEMBER AIEE

E. R. INGERSOLL
NONMEMBER AIEE

Synopsis: Saturable reactors can be used to produce narrow pulses from a periodic supply without the use of electron tubes. An increasingly popular circuit for producing pulses is known as the Melville Line. The original paper[1] describing the circuit is an excellent treatise on the theory of the circuit and includes some practical considerations. Another paper[2] presents a method of determining the number of stages in the circuit for minimum total core volume. There are limitations to the circuit which are determined by core characteristics. Also, by use of the relations developed, a designer can determine whether or not C cores can be substituted for toroidal cores to simplify winding and insulating.

THE MELVILLE LINE pulse generator produces pulses from a periodic supply. For most general applications, a sine-wave supply is selected to describe the circuit operation. However, any periodic wave shape may be used.

The upper power level of the generator is dependent upon the insulation and cooling techniques applied. The lower power level is limited by core characteristics.

The pulse-repetition rate is a multiple of the input frequency. The two basic circuits described here are for single-cycle and bi-cycle operation, and are defined by the number of pulses produced per cycle of the supply.

Nomenclature

A = effective core cross-sectional area, cm² (centimeters)
B_s = flux density at saturation, gausses
C_n = capacitance of nth-stage capacitor, farads
D_n = ratio of pulse width to pulse period for nth stage, duty cycles
e_{c1} = instantaneous voltage across C_1, volts
e_{L1} = instantaneous voltage across L_1, volts
E_0 = peak input voltage, volts
H_c = coercive force, oersteds
I_n = peak current through nth stage, amps (amperes)
I_n' = magnetizing current in nth-stage reactor, amps
J = energy in output pulse, watt-secs (watt-seconds)
l_a = air-gap total length, cm
l_i = total path length of core, cm
l_n = total path length of core in nth stage, cm
L_a = air coil inductance, henrys
L_l = leakage inductance, henrys
L_μ = self-inductance neglecting leakage, henrys
L_t = total self-inductance, henrys
L_{sn} = inductance of saturated reactor nth stage, henrys
$L_{\mu n}$ = inductance of unsaturated reactor nth stage, henrys
N_n = number of turns nth stage
R = resistance of load, ohms
μ_e = effective unsaturated permeability of core assembly
μ_i = effective unsaturated permeability of core material
σ = core volume, cm³
τ_n = half-amplitude pulse width of current in nth stage, secs
τ_R = half-amplitude pulse width of current in final stage, secs
ω_0 = input frequency

Bi-Cycle Operation

The magnetic-pulse generator in its simplest form consists of an L/C (inductance-capacitance) input circuit and an L/R/C (inductance reactance capacitance) saturable reactor circuit shown in Fig. 1. The two circuits share the same capacitor, C_1.

A sine-wave voltage, applied at the input, will produce an alternate-polarity output pulse as shown in Fig. 2. This is called bi-cycle operation, since two output pulses are produced per input cycle.

INPUT CIRCUIT

The input circuit consists of a linear L/C series circuit, resonant with the input frequency. If the circuit is connected to a sine-wave input supply at the time the line voltage is passing through zero, the voltage across the capacitor will follow the wave shape e_{c1}, as shown in Fig. 3. The equation of the voltage across capacitor C_1 for the lossless case is as follows:

$$e_{c_1}=\frac{E_0}{2}(\sin \omega_0 t-\omega_0 t\cos \omega_0 t) \tag{1}$$

A voltage peak will appear across C_1 at time: π/ω_0, $2\pi/\omega_0$, $3\pi/\omega_0$, $4\pi/\omega_0$, etc.

If allowed to continue, the peak capacitor voltage will approach a limit of Q times the applied voltage peak E_0. In bi-cycle operation, the wave shape e_{c1} is interrupted at π/ω_0 by the saturable-reactor circuit the explanation of which follows.

SATURABLE REACTOR CIRCUIT

The saturable-reactor circuit consists of an L-R-C series circuit in which L_1 is a saturable reactor. The unsaturated inductance of the saturable reactor is shown as L_{u1} in Fig. 4, and the saturated inductance as L_{s1}. The closing of switch SW symbolizes the saturation of the saturable reactor L_1.

The unsaturated reactance of L_1 is required to be much greater than the reactance of C_1.

$$\omega_0 L_{u1} >> \frac{1}{\omega_0 C_1} \tag{2}$$

The symbol >> in this paper indicates "is greater than 20." If the condition of equation 2 is fulfilled, the voltage across L_1 is approximately equal to the voltage of capacitor C_1.

$$e_{L1}\approx e_{c1} \tag{3}$$

The flux density of L_1 can be determined by the integral form of Faraday's Law

$$B=-\frac{10^8}{NA}\int e_{L1}dt \tag{4}$$

For bi-cycle operation, L_1 is required to saturate twice per cycle. At the time of saturation, the core flux density of L_1 equals B_s. During the interval 0 to π/ω_0 the total change in flux density is $2B_s$ as shown in Fig. 5. When equations 1 and 3 are used with equation 4

$$2B_s=-\frac{10^8}{NA}\int_0^{\pi/\omega_0}\frac{E_0}{2}(\sin \omega_0 t-\omega_0 t\cos \omega_0 t)dt$$

$$|B_s|=\frac{E_0 10^8}{\omega_0 NA} \tag{5}$$

The voltage, flux density, and hysteresis loop of L_1 are plotted in Fig. 5.

When L_1 saturates at π/ω_0, its inductance changes abruptly from L_{u1} to L_{s1} as represented by the switch SW closing in Fig. 6.

The series circuit consisting of C_1, L_{s1}, and R is designed for critical damping.

The current through the critically damped circuit follows the relation

$$i=\frac{E_0\pi t}{2L_{s1}}\epsilon^{-\frac{t}{\sqrt{L_{s11}C}}} \tag{6}$$

The current has a peak amplitude at $t=\sqrt{L_{s1}C_1}$, and reaches half peak amplitude at $t=0.23\sqrt{L_{s1}C_1}$ and $2.69\sqrt{L_{s1}C_1}$.

Paper 56-728, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted March 28, 1956; made available for printing April 30, 1956.

M. F. THOMPSON, R. R. TRAUTWEIN, and E. R. INGERSOLL are with Autonetics, a division of North American Aviation, Inc., Bellflower, Calif.

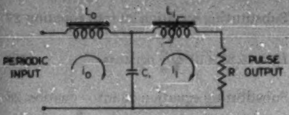


Fig. 1. Basic circuit bi-cycle operation

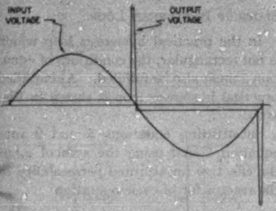


Fig. 2. Bi-cycle operation

Therefore the half-peak-amplitude width of the current pulse is

$$\tau_H = 2.69\sqrt{L_{s1}C_1} - 0.23\sqrt{L_{s1}C_1} \qquad (7)$$

or

$$\tau_H = 2.46\sqrt{L_{s1}C_1} \qquad (8)$$

The output pulse width will be arbitrarily assigned as the half-amplitude width given in equation 8.

REACTOR SELF-INDUCTANCE

The self-inductance of a reactor with a fully saturated iron core or an air coil depends only on the configuration of the coil. Let this inductance equal L_a.

The self-inductance of a reactor with a very-high-permeability core is given by the following relation:

$$L_\mu = \frac{0.4\pi N^2 \mu_e A 10^{-8}}{l_i} \qquad (9)$$

As the core saturates and μ_e approaches that of the surrounding medium (1 for air), some flux lines do not link all the turns of the coil. When the core is fully saturated, equation 9 will not yield the exact inductance unless:

1. The cross-sectional area of the core and coil are equal.
2. The winding is evenly distributed along the core.
3. The core-path length is much greater than the build-up or strip width of the core.

These three conditions cannot be fulfilled in practical reactor design, since they result in:

1. No space for insulation between core and coil.
2. Large outside dimensions of the core, even for a small cross-sectional area.

The difference between the iron-core inductance of equation 9 and the total self-inductance is called the leakage inductance and is related as follows:

$$L_t = L_\mu + L_l \qquad (10)$$

When the core is fully saturated, the total self-inductance equals the air inductance:

$$L_s = L_a = L_\mu + L_l \qquad (11)$$

Substituting equation 9 into equation 11, with a permeability of 1

$$L_l = L_a - \frac{0.4\pi N^2 A 10^{-8}}{l_i} \qquad (12)$$

Relations for L_a are available in tables for many types of configurations. Substituting equation 12 into equation 11 for permeability equal to μ

$$L_t = (\mu_e - 1)\frac{0.4\pi N^2 A 10^{-8}}{l_i} + L_a \text{ (henrys)} \qquad (13)$$

The curves of these parameters are plotted in Fig. 7.

DUTY CYCLE OF BASIC CIRCUIT

The following development is based on an ideal saturable reactor with a hysteresis loop shown in Fig. 8(A). This hysteresis loop has no coercive force, and infinite unsaturated permeability. The leakage inductance of the saturable reactor is assumed to be zero.

A measure of the ability of the pulse generator to form narrow pulses can be expressed by the duty cycle D.

$$D = \frac{\text{pulse width}}{\text{pulse period}} \qquad (14)$$

For bi-cycle operation

$$\text{pulse period} = \frac{\pi}{\omega_0} \qquad (15)$$

Substituting equations 8 and 15 into equation 14

$$D = \frac{\omega_0}{\pi} 2.46\sqrt{L_{s1}C_1} \qquad (16)$$

Since the leakage inductance is zero

$$L_l = 0 \qquad (17)$$

From equation 12

$$L_{s1} = L_a \approx \frac{0.4\pi N^2 A 10^{-8}}{l_i} \qquad (18)$$

Substituting equation 5 into equation 18

$$L_{s1} = \frac{0.4\pi}{Al}\left(\frac{E_0}{B_s\omega_0}\right)^2 10^8 \qquad (19)$$

Let

$$Al = \sigma \qquad (20)$$

Substituting equations 19 and 20 into equation 16

$$D = \frac{2.46 E_0 10^4}{\pi B_s}\sqrt{\frac{0.4\pi C_1}{\sigma}} \qquad (21)$$

The total energy in each pulse is as follows

$$J = \frac{C_1 E_0^2 \pi^2}{8} \text{ watt-secs} \qquad (22)$$

Substituting equation 22 into equation 21

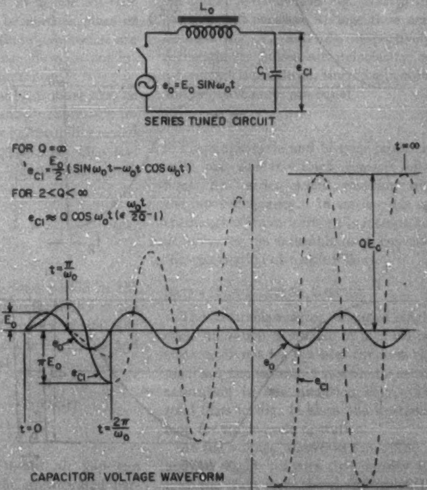

L0
$e_0 = E_0 \sin\omega_0 t$
C1
e_{C1}
SERIES TUNED CIRCUIT
FOR Q = ∞
$e_{C1} = \frac{E_0}{2}(\sin\omega_0 t - \omega_0 t \cos\omega_0 t)$
FOR 2 < Q < ∞
$e_{C1} \approx Q \cos\omega_0 t (\epsilon^{\frac{\omega_0 t}{2Q}} - 1)$
$t = \frac{\pi}{\omega_0}$
E_0
e_0
πE_0
e_{C1}
t = 0
$t = \frac{2\pi}{\omega_0}$
t = ∞
QE_0
e_0
e_{C1}
CAPACITOR VOLTAGE WAVEFORM

Fig. 3. Input circuit

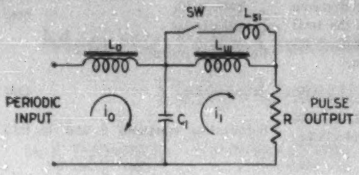

SW
L_{S1}
L_0
L_{U1}
PERIODIC INPUT
i_0
C_1
i_1
R
PULSE OUTPUT

Fig. 4. Equivalent circuit bi-cycle operation

$$D=\frac{7.93\times10^{3}}{B_s}\sqrt{\frac{J}{\sigma}} \tag{23}$$

Equation 23 shows that the output pulse width may be decreased by increasing the saturation flux density of the core material or the volume of core material. The pulse may also be narrowed by decreasing the energy content of the pulse, but this is usually a fixed design requirement that cannot be altered.

WIDTH OF HYSTERESIS LOOP

Design restrictions in a basic circuit, caused by the width of the hysteresis loop, are described in this section, and those caused by the slope of the hysteresis loop in the following section.

If a rectangular hysteresis loop is considered with a coercive force H_c and saturation flux density B_s as shown in Fig. 8(B) the magnetizing current equals

$$I'=\frac{H_c l}{0.4\pi N}\text{ amps} \tag{24}$$

For obtaining a pulse output, the peak current I must be greater than the magnetizing current, or

$$I>>I' \tag{25}$$

For critical damping and bi-cycle operation

$$I=\frac{\pi E_0}{2\epsilon}\sqrt{\frac{C_1}{L_{s1}}}\text{ amps} \tag{26}$$

Again, when the ideal reactor is considered, with no leakage inductance but with a rectangular hysteresis loop, and equation 18 is substituted into equation 26, and equations 24 and 26 into equation 25

$$H_c<<\frac{\pi E_0 10^4}{2\epsilon}\sqrt{\frac{C_1 0.4\pi}{\sigma}}\text{ oersteds} \tag{27}$$

Substituting equation 22 into equation 27

$$H_c<<5.83\times10^{3}\sqrt{\frac{J}{\sigma}}\text{ oersteds} \tag{28}$$

Substituting equation 23 into equation 28

$$D>>1.36\,\frac{H_c}{B_s}\text{ oersteds} \tag{29}$$

SLOPE OF HYSTERESIS LOOP

In the practical hysteresis loop which is not rectangular, the conditions of equation 2 must also be fulfilled. A simplified practical hysteresis loop is shown in Fig. 8(C).

Substituting equations 5 and 9 into equation 2 and using the symbol μ_e as the effective unsaturated permeability of the reactor for bi-cycle operation

$$\mu_e>>\frac{B_s^2\sigma10^{-8}}{E_0^2C_1 0.4\pi} \tag{30}$$

Substituting equation 22 into equation 30

$$\mu_e>>9.82\times10^{-9}\,\frac{B_s^2\sigma}{J} \tag{31}$$

Substituting equation 23 into equation 31

$$D^2>>\frac{0.617}{\mu_e} \tag{32}$$

Equation 32 shows that the ultimate pulse sharpening ability is limited by the unsaturated permeability. Equation 23

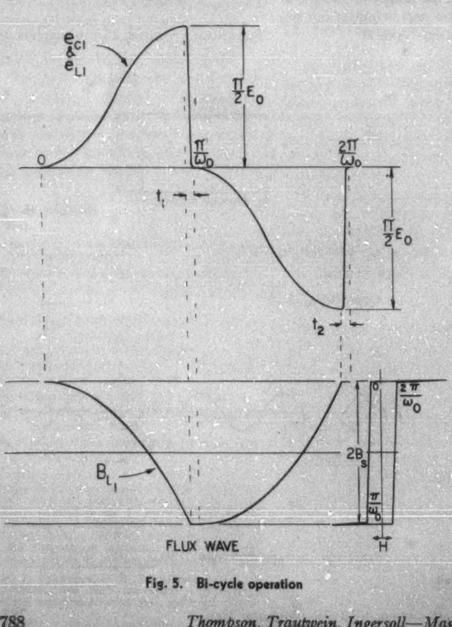


Fig. 5. Bi-cycle operation

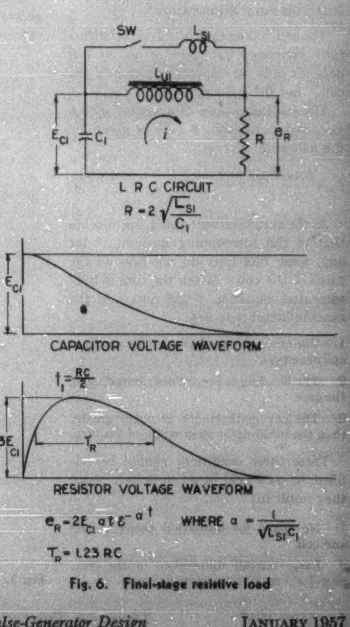

SW
L_{S1}
L_{U1}
E_{C1}
C_1
i
R
e_R
L R C CIRCUIT
$R=2\sqrt{\frac{L_{S1}}{C_1}}$
E_{C1}
CAPACITOR VOLTAGE WAVEFORM
$t_1=\frac{RC}{2}$
$0.73E_{C1}$
T_R
RESISTOR VOLTAGE WAVEFORM
$e_R=2E_{C1}\,\alpha t\,\epsilon^{-\alpha t}$ WHERE $\alpha=\frac{1}{\sqrt{L_{S1}C_1}}$
$T_R=1.23\,RC$

Fig. 6. Final-stage resistive load

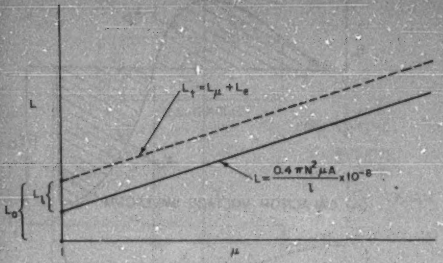


Fig. 7 (left). Reactor self-inductance

L_μ = iron core inductance neglecting leakage
L_a = air core inductance
L_e = leakage inductance

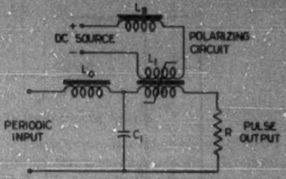


Fig. 9. Basic circuit single-cycle operation

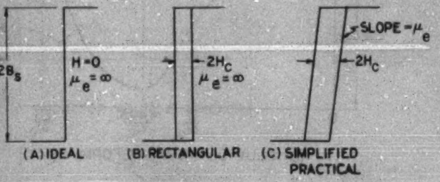
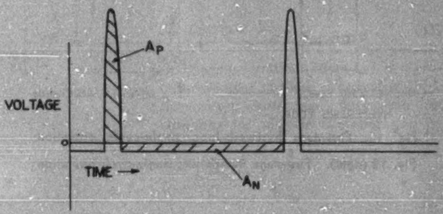


Fig. 8. Hysteresis loops

Fig. 10 (below). Voltage pulse output wave form



must be limited by equations 29 and 32 as follows

$$D=\frac{7.93\times10^3}{B_s}\sqrt{\frac{J}{\sigma}}$$

$$D>>1.36\frac{H_c}{B_s}$$

$$D^3>>\frac{0.617}{\mu_e} \tag{33}$$

"C" Core Considerations

Uncut toroidal cores with grain-oriented 50-per-cent nickel-iron material approach the rectangular hysteresis loop more closely than cut cores or silicon-iron cores. However, there are production disadvantages to the uncut toroid design that can be simplified by cut *C* cores or by stacked laminations. The stacked laminations and *C* cores permit winding of the coils separately and addition of the cores later. This advantage is particularly evident in the high-power pulse generator where high voltages are encountered, and where many turns are required in the earlier stages.

Stacked laminations have good electrical properties because of a small air gap, but cut *C* cores afford a minimum assembly time.

The rectangular hysteresis-loop characteristics of highly grain-oriented nickel-iron core material are impaired when assembled as stacked laminations or cut *C* cores. This is caused by the strain-sensitive nature of the highly grain-oriented material. However, where strain is encountered silicon-iron grain-oriented core material can be used during assembly without damaging the characteristics.

Although the Hipersil hysteresis loop is not rectangular, equation 31 states that for high pulse energies the lack of rectangularity may be permitted.

Equation 33 is a convenient arrangement of requirements to decide whether or not *C* cores may be used in place of toroids. However, these equations are only approximate, since they do not include the leakage inductance.

The effect of the air gap must also be considered. A well-known expression relates the effective core permeability μ_e and the iron permeability μ_i with the total length of the air gap l_a and the iron path length l_i.

$$\mu_e=\frac{\mu_i}{\left(\mu_i\frac{l_a}{l_i}+1\right)} \tag{34}$$

During the unsaturated period of the saturable reactor, the permeability of the core can be in the category of 100,000. If the core path length is 10 inches and the air gap is 1/10 mil

$$\mu_e=\frac{100{,}000}{\left(100{,}000\frac{0.0001}{10}+1\right)}=50{,}000$$

Thus, even a 1/10-mil air gap can seriously influence the unsaturated permeability.

Single-Cycle Operation

The bi-cycle circuit shown in Fig. 1 may be changed to single-cycle operation by applying a bias current through a bias winding and choke as shown in Fig. 9. Single-cycle operation produces unipolarity pulses at a repetition rate of 1 per cycle of the input supply. The output pulses are shown in Fig. 10 with the positive and negative voltage time area denoted by A_P and A_N respectively. Although the pulses are unipolarity, no d-c output is produced, since the positive and negative areas are equal.

Input Circuit

The single-cycle and bi-cycle input circuits are similar. Each consists of a linear L/C series circuit, resonant with the input frequency. In single-cycle operation, the wave form e_{c1} is interrupted at $2\pi/\omega_0$ by the saturable-reactor circuit, the explanation of which follows.

Saturable-Reactor Circuit

The saturable-reactor circuit of single-cycle operation is similar to bi-cycle operation, except for the bias circuit which polarizes the saturable reactor to permit saturation in one direction only. The solid lines in Fig. 11 show the hysteresis loop for single-cycle operation.

As the voltage, shown in Fig. 12(A), is applied to L_1, its core flux density follows the wave form shown in Fig. 12(B), according to Faraday's law, equation 4.

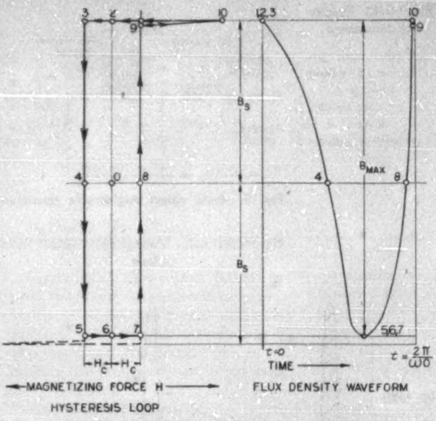
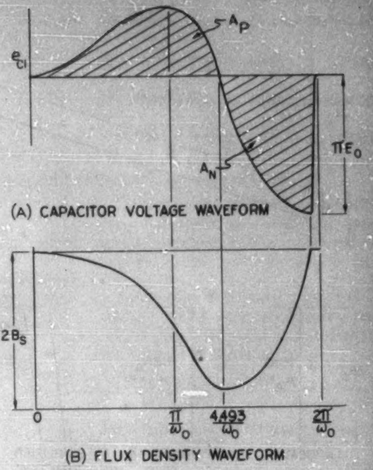


Fig. 11. First-stage hysteresis-loop single-cycle operation

Fig. 12 (right). First-stage flux-density single-cycle operation



At time $2\pi/\omega_0$ the reactor saturates and the capacitor discharges through L_1 and R. This capacitor discharge is shown in Fig. 12(A) in the vicinity of time $2\pi/\omega_0$. A finite discharge time is required, but for simplicity equation 36 assumes a zero discharge time. With this modification, the voltage wave form crosses the zero voltage axis at time $4.493/\omega_0$ and $2\pi/\omega_0$. Since the positive and negative voltage time areas are equal, the integral limits in equation 35 may be either 0 to $4.493/\omega_0$ or $4.493/\omega_0$ to $2\pi/\omega_0$. Either set of limits will give the same maximum flux density. This flux density wave form is repeated again in Fig. 11 to show its relation to the hysteresis loop.

Applying equation 4

$$2B_s = -\frac{10^8}{NA}\int_0^{4.493/\omega_0}\frac{E_0}{2}(\sin\omega_0 t - \omega_0 t\cos\omega_0 t)dt \qquad (35)$$

$$B_s = \frac{1.7E_0}{\omega_0 NA}10^8 \qquad (36)$$

Following the same manipulation as was employed under the headings "Duty Cycle of Basic Circuit," "Width of Hysteresis Loop," and "Slope of Hysteresis Loop"

$$D = \frac{3.36\times10^3}{B_s}\sqrt{\frac{J}{\sigma}}$$

$$D >> 0.574\frac{H_c}{B_s}$$

$$D^2 >> \frac{0.154}{\mu_e} \qquad (37)$$

Additional Stages

The ultimate pulse-sharpening ability of a one-stage magnetic-pulse generator is limited by the relations shown in equation 33 for bi-cycle operation, and in equation 37 for single-cycle operation. If sharpening is required beyond the limitations of these equations, additional stages must be added to the circuit.

Additional inserted stages are shown in Fig. 13. This circuit has been chosen for bi-cycle operation, to illustrate how stages are added. A bias could be added to L_1 as in Fig. 9, if single-cycle operation is desired.

All additional stages are L/C-series circuits, sharing capacitors with adjacent stages. The stage immediately following the input stage is defined as the first stage. Stages beyond the first stage are intermediate stages. The stage including the load is called the final stage. The first stage and intermediate stages are identical in operation during their saturated period, but during their unsaturated period the wave forms of the two types differ.

In all the following equations, all capacitors will be equal or

$$C_1 = C_2 = C_3 = C_n = C \qquad (38)$$

FIRST STAGE

The first stage circuit consists of C_1 and C_2 in series with L_1 as shown in Fig. 13.

For bi-cycle operation, equations 2 through 6 apply directly, while L_1 is unsaturated. However, when L_1 is saturated, the circuit becomes an L/C/C (inductance capacitance capacitance) instead of an L/R/C circuit, and equations 7 and 8 do not apply. The wave forms for the first stage circuit, while L_1 is saturated, are shown in Fig. 14

At time π/ω_1 the charge on C_1 has been completely transferred to C_2. For the linear, lossless case, the circuit would oscillate indefinitely. However, L_2 saturates at π/ω_1 and C_2 is discharged through the second stage. The discharge time of C_2 is much less than π/ω_1 as shown in Fig. 14.

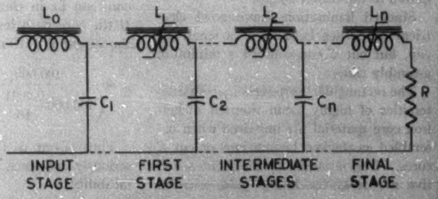


Fig. 13. General circuit

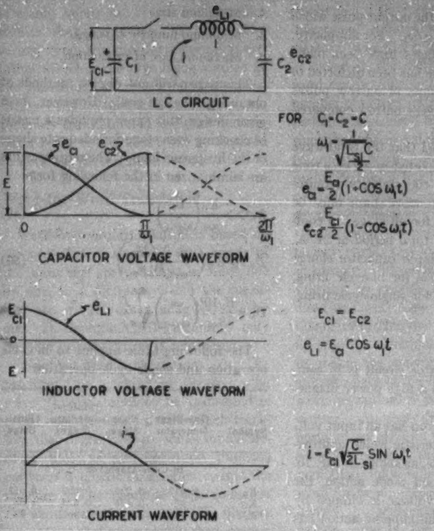

e_{L1}
E_{C1}
C_1
i
C_2
e_{C2}
L C CIRCUIT
FOR $C_1 = C_2 = C$
$\omega_1 = \frac{1}{\sqrt{\frac{L_{s1}}{2}C}}$
$e_{C1} = \frac{E_{C1}}{2}(1+\cos\omega_1 t)$
$e_{C2} = \frac{E_{C1}}{2}(1-\cos\omega_1 t)$
e_{C1}
e_{C2}
E
0
$\frac{\pi}{\omega_1}$
$\frac{2\pi}{\omega_1}$
CAPACITOR VOLTAGE WAVEFORM
$E_{C1} = E_{C2}$
$e_{L1} = E_{C1}\cos\omega_1 t$
E_{C1}
0
E
e_{L1}
INDUCTOR VOLTAGE WAVEFORM
$i = E_{C1}\sqrt{\frac{C}{2L_{s1}}}\sin\omega_1 t$
i
CURRENT WAVEFORM

Fig. 14. Operation after saturation; first and intermediate stages

The half-amplitude width of the first stage current pulse equals

$$\tau_1 = \frac{\pi}{3}\sqrt{2L_{s1}C} \tag{39}$$

A procedure similar to that described in the afore-mentioned three sections can be used to determine the first-stage duty-cycle restrictions for bi-cycle operation.

$$D_1 = \frac{4.76\times 10^3}{B_{s1}}\sqrt{\frac{J}{\sigma_1}}$$

$$D_1 >> 0.425\frac{H_{c1}}{B_{s1}}$$

$$D_1^2 >> \frac{0.222}{\mu_{s1}} \tag{40}$$

For single-cycle operation, the first-stage duty cycle restrictions are as follows:

$$D_1 = \frac{2.03\times 10^3}{B_{s1}}\sqrt{\frac{J}{\sigma_1}}$$

$$D_1 >> 0.180\frac{H_{c1}}{B_{s1}}$$

$$D_1^2 >> \frac{0.0555}{\mu_{s1}} \tag{41}$$

INTERMEDIATE STAGE

The intermediate-stage circuit is identical to the first-stage circuit. However, the voltage applied to the intermediate-stage saturable reactor before saturation is sinusoidal.

For bi-cycle operation

$$e_{Ln} = \frac{E_0\pi}{2}\sin\omega_{n-1}t \tag{42}$$

$$\omega_{n-1}L\mu_n >> \frac{1}{\omega_{n-1}C} \tag{43}$$

By using equations 42 and 43 in a procedure similar to those already described, the intermediate-stage duty-cycle restrictions for bi-cycle operation can be found

$$D_n = \frac{5.89\times 10^3\omega_0}{B_{s_n}\omega_{n-1}}\sqrt{\frac{J}{\sigma_n}}$$

$$D_n >> \frac{0.524\omega_0 H_{c_n}}{B_{s_n}\omega_{n-1}}$$

$$D_n^2 >> \frac{0.222}{\mu_{s_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)^2 \tag{44}$$

For single-cycle operation, the intermediate-stage duty-cycle restrictions are as follows:

$$D_n = \frac{2.95\times 10^3}{B_{s_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)\sqrt{\frac{J}{\sigma_n}}$$

$$D_n >> \frac{0.262H_{c_n}}{B_{s_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)^2$$

$$D_n^2 >> \frac{0.0555}{\mu_{s_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)^2 \tag{45}$$

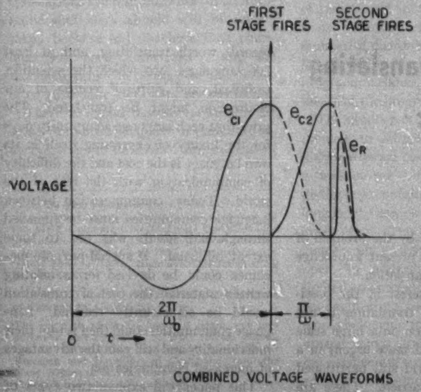


Fig. 15. Single-cycle operation

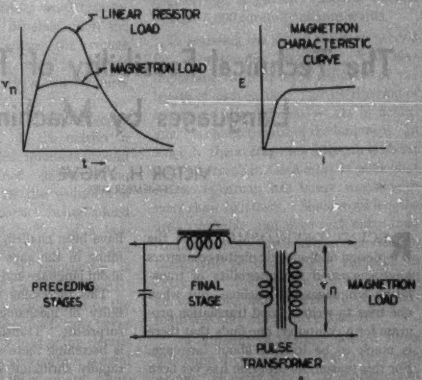

LINEAR RESISTOR LOAD
MAGNETRON LOAD
V_n
t →
MAGNETRON CHARACTERISTIC CURVE
E
I
PRECEDING STAGES
FINAL STAGE
V_n
MAGNETRON LOAD
PULSE TRANSFORMER

Fig. 16. Final-stage magnetron load

FINAL STAGE

The final-stage and intermediate-stage wave shapes are identical before saturation.

Therefore, for bi-cycle operation, equations 42 and 43 apply directly. However, after saturation, the final stage circuit becomes a critical damped L/R/C circuit, and the half-amplitude width of the current pulse equals

$$\tau_n = 2.46\sqrt{L_{s_n}C} \qquad (46)$$

Again, applying a similar procedure the final-stage duty-cycle restrictions for bi-cycle operation are determined.

$$D_n = \frac{9.79\times10^3}{B_{s_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)\sqrt{\frac{J}{\sigma_n}}$$

$$D_n >> \frac{1.68H_{c_n}}{B_{s_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)$$

$$D_n^2 >> \frac{0.617}{\mu_{e_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)^2 \qquad (47)$$

For single-cycle operation, the final-stage duty-cycle restrictions are

$$D_n = \frac{4.88\times10^3}{B_{s_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)\sqrt{\frac{J}{\sigma_n}}$$

$$D_n >> \frac{0.836H_{c_n}}{B_{s_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)$$

$$D_n^2 >> \frac{0.154}{\mu_{e_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)^2 \qquad (48)$$

Combined Wave Forms

The combined wave forms of a 2-stage single-cycle pulse generator are shown in Fig. 15. In actual circuits, the time period $\pi/\omega_1 << 2\pi/\omega_0$ and the output pulse width is much less than π/ω_1. For simplicity, the intermediate stages have been omitted. The time scale has been distorted to show the wave shape of narrow pulses which are actually very narrow compared to the input period.

It should be noted that the total wave shape must be completed in a $2\pi/\omega_0$ time interval to satisfy the stability requirement that the input wave shape must be shortened to allow for all stages to complete the cycle in a total period of $2\pi/\omega_0$. Consequently, the input capacitor charge time is less than π/ω_0 for bi-cycle firing, and less than $2\pi/\omega_0$ for single-cycle firing.

Magnetron Load

The pulse generator circuit is in common usage to drive a high-power magnetron.

A typical magnetron has an input voltage-current characteristic which approximates a biased diode, as shown in Fig. 16. As a result of biased diode action, the final-stage output voltage is clipped by the magnetron. This clipping action has a tendency to square the magnetron output pulse.

Design Procedure

A magnetic-pulse generator may be designed to meet one or more of the following design goals in accordance with its particular application:

1. Minimum core volume.
2. Maximum efficiency.
3. Minimum weight.
4. Minimum size.
5. Minimum number of stages.
6. Maximum ease of production.

This paper will not discuss methods of obtaining a design goal. However, for a given design, this paper provides a means of checking each stage for adequate shape of the hysteresis loop. The requirements are summarized in the following form

$$D_n^2 >> \frac{K_u}{\mu_{e_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)^2 \qquad (49)$$

$$D_n >> K_H\left(\frac{\omega_0}{\omega_{n-1}}\right)\left(\frac{H_{c_n}}{B_{s_n}}\right) \qquad (30)$$

$$D_n = \frac{K_J10^3}{B_{s_n}}\left(\frac{\omega_0}{\omega_{n-1}}\right)\sqrt{\frac{J}{\sigma_n}} \qquad (51)$$

The following table applies to bi-cycle operation and single-cycle operation:

Symbol	One-Stage Generator	First Stage	Intermediate Stages	Final Stage
Bi-Cycle Operation				
K_u	0.617	0.222	0.222	0.617
K_H	1.36	0.425	0.524	1.68
K_J	7.93	4.76	5.80	9.79
Single-Cycle Operation				
K_u	0.154	0.0555	0.0555	0.154
K_H	0.574	0.180	0.262	0.836
K_J	3.36	2.03	2.95	4.88

References

1. THE USE OF SATURABLE REACTORS AS DISCHARGE DEVICES FOR PULSE GENERATORS, W. S. Melville. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 98, pt. III, 1951, pp. 185–207.

2. ECONOMIC DESIGN OF SATURATING REACTOR MAGNETIC PULSERS, R. A. Mathias, E. M. Williams. *AIEE Transactions (Electrical Engineering)*, vol. 74, pt. I, May 1955, pp. 169–71.

The Technical Feasibility of Translating Languages by Machine

VICTOR H. YNGVE
NONMEMBER AIEE

RECENT DEVELOPMENTS in the design of electronic digital computers have suggested the possibility of translating languages by machine, but when one tries to write a good translation program for a computer, one finds that there is much to be learned about language. For this reason, no program has yet been written by which a computer could translate usefully; but a number of people have been challenged by the problem of filling in the gaps in present knowledge about language and translation.[1,2]

This increasing interest in the possibility of mechanical translation is not surprising. Translation on a large scale is becoming more and more urgent in a rapidly shrinking world still partitioned by many language barriers. There are at least 50 languages that possess written records worth translating, and at least 200 languages into which the scientific, technical, and cultural records of our civilization might be translated. The price that each language community pays for the luxury of expressing itself in its own language is the cost and the difficulty of communication with the rest of the world. Today, communication between language communities must be funneled through individuals who are to some extent bilingual. If special-purpose machines could be devised for translating written material, the cost of translation would be significantly reduced. Language communities could then retain their individuality and still gain the advantages of increased communication.

This paper will explore two types of mechanical translation. The first, which can be called word-for-word translation, is within immediate reach, but promises results that are crude, at best. The second type, sentence-for-sentence translation, which is the type being studied at the Research Laboratory of Electronics, Massachussets Institute of Technology, promises translation of a higher quality, but a great deal of effort will be required to work out the linguistic details.

Word-for-Word Translation

Word-for-word translation is based on the use of a dictionary. The words of the input text are looked up in order, one by one, in a dictionary, and the meanings are written down in order, one by one. In determining the meanings, only one text word at a time is considered.

Word-for-word translations can be produced by machine in the following way: The material to be translated is typed on a keyboard device which transfers the characters to some input medium such as punched tape, punched cards, or magnetic tape for entry into the machine. The machine then compares the words one at a time with the dictionary entries that are stored in the memory. When each input word is located in the memory, the dictionary entry stored there provides a meaning that consists of an equivalent word or words in the other language. The equivalents found in the dictionary are then sent one at a time to an output printer that prints a word-for-word translation.

If a machine that can produce rough translations on a production basis is wanted as soon as possible, the word-for-word machine is the favored type because it is simple and can be implemented now. If translations better than word-for-word translations are wanted, work on an automatic dictionary should still be undertaken because any machine that translates will need a dictionary, and the technical problems arising in the design of an automatic dictionary are not trivial.

There are good reasons for taking the word, delimited by spaces or punctuation, as the item to be looked up, instead of some smaller unit. It has been suggested that if the roots of words and their inflectional endings were stored in the memory separately, much space would be saved in the dictionary. If a given word can be inflected in five ways, the number of entries in the dictionary could, perhaps, be reduced by nearly four-fifths, since the inflectional endings are common to many words. As another space-saving device, it has been suggested that compound words, such as *blackboard, counterelectromotive, Koordinatenanfangspunkt*, could be separated into their component parts, and each part looked up separately. Suggestions such as these for splitting up words have at least two disadvantages. One is that the separation of a word into parts before it is looked up requires a method for deciding where to make the separation. Part of the storage space saved has to be used for a more complicated program or look-up routine. Another disadvantage is that parts of words and inflectional endings have, in general, more different meanings than fully inflected words. Strict word-for-word translation completely avoids the segmentation problems and avoids some of the multiple meanings of parts of words—inflectional endings, in particular.

In the next sections, word-for-word translation will be considered in detail. The requirements of the task will be compared with the capabilities of present-day, general-purpose digital computers and of special-purpose machines that might be constructed for translation; then the usefulness of the product will be assessed.

MACHINE REQUIREMENTS: SIZE

The size of the memory of a machine is of importance in word-for-word translations. Memory size depends upon the number of dictionary entries, the average length of each entry, and, in stored-program machines, on the storage requirements of the program.

The number of words in a language is difficult to estimate. If the entries in dictionaries are counted, numbers around a million are obtained. However, no dictionary contains all the words of a language; new words are being added every day. The vocabulary of a language is not a closed system; it is always open to additions. If translating is limited to one field of knowledge (electrical engineering, for example) most of the words belonging to other fields may never appear, and a glossary containing electrical engineering terms and certain general scientific terms might be adequate. But, like the vocabulary of an entire language, the vocabulary of a specific field is not closed; and an author of a paper in a special field often uses terms from other fields.

A glossary should contain the words most frequently used in a specific field. Since it is virtually impossible to list all the words of a language, or even of a specific field, new words will have to be added from time to time as they are needed. It is feasible to start with a rather small list of words and add new ones where they are first encountered or after they have appeared a given number of times. In this way, the glossary is always an approximation to a list of the words of highest frequency in the specific field.

The most important parameter to be considered in evaluating a glossary is not the number of words that it contains, but some parameter connected with the number of running words of a text that can be found in the glossary. The coverage of a glossary with respect to a certain test can be defined as the ratio of the number of running words of the text that can be found in the glossary to one more than the number of words that cannot be found. It is the average length of uninterrupted runs of text words that can be found in the glossary. The coverage of an English glossary with respect to a Russian article is zero. The coverage of a glossary containing the 50 or 60 most frequently used English words is about one, because these words make up about half of the running words in a text. The coverage of a glossary with respect to an article is numerically equal to the number of words in the article if the glossary contains them all. To have a reasonably high coverage with respect to a field such as electrical engineering, a glossary should contain at least 5,000 carefully chosen words.

In assessing the size of memory needed for an automatic dictionary, it is also important to form some estimate of the length of each entry. The distribution of lengths of the different words in a language is characteristic of the language. In English, the average word length is about 7 or 8 letters; in German it is somewhat longer. German has many words with more than 20 letters. Since word lengths are so variable, and since each entry may contain one or more words as equivalents, there will be considerable variation in entry size. Perhaps the average size will be something like 20 letters; the maximum size may run to more than 100.

The amount of storage needed for a minimum dictionary of 5,000 entries with an average entry size of 20 characters of

---

Paper 56-928, recommended by the AIEE Communication Theory Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted June 11, 1956; made available for printing July 19, 1956.

VICTOR H. YNGVE is with Massachusetts Institute of Technology, Cambridge, Mass.

The work on which the paper is based was supported in part by the U. S. Army (Signal Corps), the Air Force (Office of Scientific Research, Air Research and Development Command), and the Navy (Office of Naval Research); and in part by the National Science Foundation.

5 binary digits each is 500,000 binary digits.

Now that an estimate of the memory capacity requirements has been made, the adequacy of present-day machines can be examined. Many general-purpose digital computers have three kinds of memory: magnetic core, magnetic drum, and magnetic tape. A storage capacity of 500,000 binary digits is, by present standards, large for a magnetic core memory. Even many magnetic drum machines do not have this capacity. One is forced to the conclusion that magnetic-tape storage must be used with most present-day general-purpose computers if an adequate dictionary is to be stored.

MACHINE REQUIREMENTS: SPEED AND ECONOMY

While the ability of a machine to make word-for-word translations is determined primarily by the size of the glossary that it can store, its usefulness for producing translations economically depends on a number of factors. These include the cost of operating the machine, the access time of the memory, the register length, the type of commands available and their execution times, and the way in which the program is written.

In present-day machines the memories are organized by address. This means that the memory contains a number of pigeon holes or memory locations, each one designated by a number. The machine is arranged so that ready access is obtained to each memory location when it is referred to by its address. The average time that it takes to refer to a given address and extract the contents of the memory from that position is called the access time. For core memories the access times are very short, in the neighborhood of 10 to 20 microseconds. For drum and tape memories, the concept of access time is not quite as appropriate, since the actual time required to find something at a given address varies greatly, being dependent on the time it takes the drum or tape to move to that address. Average access times for drums are around 10 milliseconds.

In a dictionary look-up program, the access time is not the only factor in determining how quickly a given word can be found; although the word is known, its address is not. The binary representation of the word itself cannot be used as the address, because the spelling of words is highly redundant. For a 5,000-word memory, the address would need to be only 13 binary digits long; but the spelling of a word might require well over 100 binary digits.

Various programming methods have been devised for searching a memory when the address is unknown. Clearly, it is not economical of time to start at the beginning and examine each entry in turn, through the whole dictionary, until the right entry is found. An improvement consists of using the thumb index device, whereby the search is divided into two steps. The initial one or two letters of the word are found in the index part of the dictionary. The index entry gives the position in the main list at which the search should begin. It is possible to extend the thumb index idea and divide the search into three or more steps. If this procedure is carried to the extreme, the bifurcation, or "20-questions," technique is reached. Each step of the search chooses one out of two approximately equal sections of the dictionary. If the division is exactly into two each time, one out of 2^n entries can be found in n trials. Another possibility is to try to obtain from the spelling of the word some function that specifies, as closely as possible, the address of the word. With this scheme, surprisingly good results can be obtained.

It is also possible to take advantage of the vocabulary statistics. The probability p_r that the rth word from the vocabulary will appear next in a text is given by the Zipf law[3] as $p_r = P\,r^{-1}$, where r is the rank of the word when words are ranked in order of decreasing frequency, and P is a constant with a value of approximately 1/10. A more accurate law has been derived by Mandelbrot.[4] It is $p_r = P\,(r+m)^{-B}$, where B is a constant in the neighborhood of one, and m is another constant. The constants P, m, and B are characteristic of different languages and, to some extent, of different authors and styles of discourse. According to these two laws, as few as 50 or 60 different words make up about half of the total words in a text. If these 50 or 60 words are placed in a core memory with effectively instantaneous access, the over-all access time can be reduced by a factor of 2.

For a memory such as magnetic tape, the access time to a given entry depends on its position along the tape in relation to the reading heads. The most effective way to use a magnetic-tape memory is to look up the words in batches, not one at a time. Each batch of words is alphabetized in the fast memory of the machine. All of the words in a batch can be looked up in one pass of the tape, if the words on the tape are also alphabetized. The average rate of access can be greatly increased by this method at the expense of some delay in translation and some storage and manipulation requirements in a fast memory. The concept of random access time is almost meaningless when words are looked up in batches. The appropriate parameters of the system are the average rate of information flow through the machine, and the average delay.

Present-day digital computers have been designed to handle arithmetic at very fast rates. A dictionary search program for one of these machines usually involves a great deal of arithmetic, not because it is required by the problem but because the machine does it so well. A modern fast computer would not be as economical as a machine specifically designed for translation. Such a machine should be capable of handling entries of variable length and designed specifically for searching the memory when the address is not known. It should be specially designed for matching sequences of binary digits, and determining whether the numerical value of one sequence is greater or less than that of another.

Memory research is progressing rapidly. In another few years, high-capacity magnetic-disk memories, special magnetic tape or drum arrangements, or various types of photographic memories will probably be developed to the point where they will be available for translating machines. It may not be necessary to resort to word-splitting techniques in order to store a high-coverage glossary in a machine that is economical to use.

The simple word-for-word translation schemes that have been discussed have the advantage that they can be implemented now. To decide whether or not their implementation is worth while, the potential usefulness of the translations should be considered.

USEFULNESS OF OUTPUT

If translation, from German to English, for example, is considered as the selection of one out of all possible sequences of English words for each German sentence, a word-for-word translation is a good first approximation. English sentences average about 20 words. Assuming a vocabulary of 10^5 words, there are 10^{100} possible sequences of this length. A word-for-word translation may supply an average of about two meanings for each German word. If the word-for-word translation is good, the correct English translation can be found by choosing the correct meaning for each word from those supplied, and then choosing the correct word order. The initial choice of one out of 10^{100} possible sequences has been

Vol.
Pt. 1
1956

reduced to one out of $2^{30} \times 201$, or about one out of 10^{34}. For each sequence left to choose from, approximately 10^{66} have been ruled out. It is thus seen that the word-for-word translation has done a Herculean job, but there is still much work for the reader before he can read the translation. He must still choose one out of 10^{34} possibilities. This he presumably does on the basis of his knowledge of English sentence structure and on the basis of his knowledge of the subject matter. The problems of multiple meaning and of word order are the two difficulties that he must face.

To make the reader's task easier, various suggestions have been made for the arrangement of the multiple choices on the page. They could be strung out in linear order, perhaps with parentheses; or the most likely meaning for each word could be put first or written in bold type; or the various meanings could be listed in columns. Another question concerns the number of meanings that should be provided for each word. Perhaps one broad meaning would be less confusing than two or three narrower meanings. Another suggestion for making the reader's task easier is called partial translation. Partial translation is based on the fact that only a few hundred articles, prepositions, conjunctions, and adverbs make up a large majority of the running words in a text and that these words, together with a few inflectional endings, have especially long multiple-meaning lists. The suggestion is that these few hundred items be given in their original language, not to confuse the reader with long lists. The reader, after a brief introduction to the grammar and sentence structure of the foreign language, would be able to understand these words and endings and the constructions in which they are used. All of the technical terminology would be translated for him.

The reader of the translation plays a vital role. In a sense it is he who makes the translation, because he alone understands the meaning. The machine only recodes the foreign language so that it can be more easily translated by the reader. Experiments have shown that word-for-word translations can be deciphered more easily if the reader is familiar with the topic that is being discussed. If he knows what the author is trying to say, he can recreate the work with very few clues. Thus, word-for-word translation could be useful for people who want to keep abreast of large amounts of foreign technical literature in their field, but it would probably not be adequate for a careful reading of any one article.

It has been suggested that rough word-for-word translations should be revised and edited by an expert. For certain purposes, revision may be advisable, but perhaps the best course would be to let the final reader make his own revision, since he is an expert in the field and can use his background knowledge in deciding what the author probably meant.

It is very important that the final reader be considered a link in the translation chain. Any improvement in the translation which speeds up the reader's comprehension is worth while, since the reader's time is expensive.

CONCLUSIONS

Word-for-word translation can be implemented now by using present-day general-purpose digital computers or, more economically, special-purpose machines. The output might be useful, but would definitely be of poor quality, probably so poor that the reader would have to be highly motivated before he would want to use it.

Sentence-for-Sentence Translation

Attention will now be given to the question: How can a machine produce something better than a word-for-word translation? The deficiencies of word-for-word translations will be examined, not from the point of view of how a reader may cope with them, but from the point of view of how a machine may do so. A diagnosis of the difficulties will be attempted, and a remedy will be proposed. This will necessitate some examination of the way in which words are combined to make sentences in a language.

ANALYSIS OF DIFFICULTIES

The major shortcomings of word-for-word translations have been grouped under multiple-meaning difficulties and word-order difficulties. It is now advantageous to reclassify them as lexical difficulties and structural difficulties.

Lexical difficulties include the multiple meanings of words like *power*, *force*, *integrate*, *potential*, and so on. It has been suggested that the use of a field glossary would restrict these words to one meaning, the technical one. A startling exception to the observation that most nouns, verbs, and adjectives have few meanings is the word *run* which, according to Webster, has 54. (The average number of meanings per word in *Webster's Collegiate Dictionary* is about 2.5.)

Structural difficulties include word order, inflectional endings, multiple meanings of most articles, prepositions, conjunctions, and adverbs. Most of the multiple-meaning difficulties in a text are structural because a majority of the running words in a text are structural words, with more than the average number of meanings. The 54 meanings of *run* are divided into four lists that are structurally different: transitive verb, intransitive verb, noun, and adjective. It is clear that a machine method of handling structural problems is highly desirable.

Diagnosis of the difficulties is easy. After all, a word is usually not ambiguous when its context is considered. In word-for-word translation, no attempt is made to translate a word in its context, but a meaning or meanings are assigned to it on the basis of its form alone. From this point of view, a word-for-word translation might be called a literal translation, whereas a translation which takes into account the context might be called a free translation. Whether a translation is literal or free really depends upon the point of view. A word-for-word translation translates words literally, but translates letters and syllables freely. Words may be translated freely if larger blocks are translated literally. In general, a literal translation will be better if it is based on larger blocks of text.

A way is needed to use the context in translating the words. Investigation has shown that using the context of a complete sentence is a great deal better than using smaller parts of a sentence, say a phrase, but is not significantly worse than using the context of two or more sentences. A sentence seems to be a natural language unit. To see how the information in a sentence context can be used for translation, more knowledge of language structure is needed.

STRUCTURE OF LANGUAGE

The most significant thing which may be said about a sentence is that it has a structure. Some hint of this is given in the grammar taught in grammar school, where words and groups of words are classified in terms of nouns, verbs, subjects, objects, and so on, and simple sentences are parsed in order to show the interrelation of their parts.

It is rather illuminating to consider some of the quantitative aspects of sentence structure. Of the 10^{100} arbitrary sequences of 20 words, perhaps only 10^{25} to 10^{30}, are actually English sentences. Practically all of the remainder are mere incoherent jumbles of words. Sentence structure thus places very strong constraints on the sequences of words. Another way of appreciating the degree of constraint imposed by sentence structure is to take a 20-word sentence from a newspaper and rearrange the words. Of the 20! different arrangements, only a few are English sentences. That there must be some structure or system underlying these constraints is apparent. A man in his lifetime would not be physically able to hear or speak more than 10^9 sentences, yet he can instantly recognize whether any given sequence of 20 words is included among the 10^{30} to 10^{50} English sentences or is left in the remaining 10^{100} ungrammatical jumbles of words. He can do this by recognizing whether or not the sequence conforms to one of the allowable sentence patterns. Recognition of sentence pattern is based primarily on formal features of the sequence and not on the meaning; for the meaningless lines by Lewis Carroll:

> 'Twas brillig, and the slithy toves
> Did gyre and gimble in the wabe:
> All mimsy were the borogoves,
> And the mome raths outgrabe.

are instantly recognized as conforming to English sentence patterns.

Constraints powerful enough to restrict the sequences of words to only those that conform to the sentence structures of a language can be represented by the following scheme. A set of structural choices is encoded into a sequence of words. The first step is to choose one of several possible sentence types. For English these might be:

sentence = 'I' + predicate a

or

sentence = 'you' + predicate b

or

sentence = subject c + predicate c

etc., or

sentence = 'it' + predicate p + 'that' clause

etc.

In this "sentence" stands for all sentences in English; "predicate a" stands for all predicates that can follow the first person, singular subject "I"; and so on. Take the example:

sentence = 'it' + predicate p + 'that' clause

This means that the sentence will start with the word 'it,' followed by a particular type of predicate, predicate p, and followed in turn by a clause introduced by the word 'that.' The particular predicate of type p and the particular 'that' clause will emerge as a result of further choices. For the predicate, the choices might be:

predicate p = verb 'to be' + adjective b

or

predicate p = verb 'to be' + prepositional phrase

or

predicate p = verb 'to be' + past participle c

etc.

The first of these might be chosen. Further choices would decide the tense of the verb, the present tense, for example. For the adjective, the choices might be:

adjective b = 'clear'

or

adjective b = 'true'

or

adjective b = 'likely'

etc.

If 'clear' is chosen, the sentence so far is: "It is clear + 'that' clause." Note that adjective b would not include adjectives like hot, red, small, and so on.

An encoding scheme of this sort can allow for certain types of recursiveness in sentence structure. For example:

It is clear that it was likely that it was true that...

This type of recursiveness is indicated by

'that' clause = 'that' + sentence

which leads back again to the beginning of the set of choices.

An encoding scheme of this sort, starting as it does with a set of choices or directions for producing a certain sentence and ending with the words of that sentence, can be thought of as a grammar of a language from the point of view of the speaker rather than of the listener. Schemes from the speaker's point of view have the advantage that each set of choices leads unambiguously to a unique sequence of words. On the other hand, a given sequence of words may be ambiguously represented by two or more different sets of choices. As an example of this, consider that

predicate a = 'am here now'

or

predicate a = 'see it now'

etc., and

predicate b = 'are here now'

or

predicate b = 'see it now'

etc.

Then 'see it now' can be produced ambiguously as a first-person singular or a second-person predicate. A grammar from the listener's point of view could resolve this ambiguity only by consideration of the context 'I' or 'you.' Similarly, words can have multiple meanings from a literal point of view, but be unambiguous if their role in the sentence structure is known. As an example, consider this English sentence:

The (Der) man showed the (dem) boy the (das) picture.

and this part of a German sentence:

Der (the) Mann, der (who) der (of the) Tat fähig ist,...

In this way, considerations of sentence structure can resolve many of the multiple-meaning problems.

Usually each possibility of choice of structure in a language has a meaning assigned to it. A distinction has been made between the "structural" meaning and the "lexical" meaning. The two sentences "The chair is on the rug." and "The rug is on the chair." have different meanings because the words appear at different positions in the structure. On the other hand, the difference in meaning between "The chair is on the rug." and "The chair is on the floor." is a lexical difference. Lexical meaning tells what sentences talk about; structural meaning tells what they say about it.

Just as different languages are different lexically, they can also be different structurally. Languages that are historically related are usually similar in structure. Even if two languages have certain structures in common, it may be that they assign different meanings to these structures. Languages may also differ in what aspects of the meaning are carried structurally and what aspects are carried lexically. The distinction between subject and object in a sentence is indicated structurally by word order in English, but in German it may be indicated by the lexical difference between *der* and *den*. Similarly the meaning that is assigned lexically to the word *if* in English, is sometimes indicated structurally by word order in German.

TRANSLATION OF STRUCTURE

If translation by machine is to be done on a sentence-by-sentence basis, grammars are needed for each language. The use of sentence structure to resolve word order and multiple-meaning problems and to translate the structural meanings correctly must be based on adequate grammars. These grammars should consist of descriptions of each language which are complete enough to make the distinctions in meaning necessary for translation. A description of this sort for even one language has not yet been provided, because the concepts involved are just now beginning to be understood and because languages are extremely complex in structure. Once two of these descriptions are available, they can be set side by side, and the equivalences stated. It must be decided which sentences in language *A* and language *B* are equivalent. The underlying assumption is that it is possible to find such correspondences. Of course, this assumption is not strictly valid, for no two languages are completely translatable. It is probably very nearly valid for related languages such as German, English, French, and Russian, and for scientific discourse, or other types of careful exposition.

If a description of the structure of two languages and a statement of the equivalences in meaning between them are given, only one more thing is needed for a mechanized translation procedure: a mechanical method of recognizing the structure of a sentence from the sequence of words composing the sentence. The precise form of this recognition routine must await the description of the structures to be recognized. However, the process must operate from the point of view of the person who reads or hears a sentence. It must tentatively recognize small groups of words, place them in a tentative structure, and so proceed, each time eliminating certain possibilities. Once the structure of the sentence to be translated has been recognized, it can be translated into an equivalent structure in the other language, and the correct meanings can be chosen from the multiple meanings of the words.

RESULTS EXPECTED

The translations produced by a routine that translates on a sentence-for-sentence basis will be vastly better than the output of a word-for-word translating machine. Most output sentences will be grammatically correct. There will be very few word-order problems remaining. Many of the multiple-meaning difficulties will be solved, particularly those involving the most frequent words of the language and the very frequent inflectional endings. There will be a partial solution to multiple-meaning problems such as the different meanings of *can* which are distinguished on the basis that one meaning is connected with a verb, one with an auxiliary verb, and one with a noun. Some multiple-meaning problems will remain. Of these, some can be resolved by judicious use of a field glossary or by related techniques. The rest must be resolved by the good sense and intelligence of the reader of the translation.

Conclusions

The state of the art of mechanical translation may be summed up in the following manner:

1. Word-for-word translations can be made now on high-speed, general-purpose, digital computers.

2. Word-for-word translations could be made more economically by means of special-purpose machines built with existing technology.

3. Word-for-word translations promise to be considerably cheaper than man-made translations.

4. Word-for-word translations are very crude, but may be useful when more accurate translations are not worth the additional cost.

5. If something better than a word-for-word translation is desired, the best procedure would be to take into consideration the sentence structure in designing a translation routine.

6. Providing translation routines on a sentence-for-sentence or structural basis requires a considerable amount of detailed linguistic work.

References

1. MACHINE TRANSLATION OF LANGUAGES (book), W. N. Locke, A. D. Booth, editors. Technology Press, Massachusetts Institute of Technology, Cambridge, Mass.; John Wiley & Sons, Inc., New York, N. Y.; Chapman and Hall, Ltd., London, England, 1955.

2. MECHANICAL TRANSLATION, Massachusetts Institute of Technology, Cambridge, Mass., vols. 1, 2, 1954 and 1955.

3. HUMAN BEHAVIOR AND THE PRINCIPLE OF LEAST EFFORT (book), G. K. Zipf. Addison-Wesley Press, Cambridge, Mass., 1949.

4. SIMPLE GAMES OF STRATEGY OCCURRING IN COMMUNICATION THROUGH NATURAL LANGUAGES, Benoit Mandelbrot. *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, New York, N. Y., Mar. 1954, pp. 124–37.

Integrated s-Plane Synthesis Using 2-Way Root Locus

JOHN ZABORSZKY
MEMBER AIEE

IN GENERAL, the design of automatic feedback control devices starts with two sets of data: 1. The specifications of the closed-loop performance such as the steady-state errors or error constants, minimum speed and maximum overshoot of the transient, bandwidth, etc., and 2. The known characteristics of physically existing components, combined into a plant which must become a part of the open loop of the control system.

It should be noted that group 1 defines certain characteristics of the closed loop, and group 2 certain characteristics of the open loop. This is unfortunate, mainly since the correlation between closed- and open-loop characteristics is shrouded in the difficulties of factoring polynomials and computing transients, and in the labor involved. The result is a certain amount of indirectness and lack of symmetry in the techniques which have been developed for the synthesis of automatic control systems.

Nomenclature

$c(t)$ or $C(s)$ = controlled output, Fig. 1
$r(t)$ or $R(s)$ = controlling input
$e(t)$ or $E(s)$ = error
$G(s) = G_1(s)G_2(s)$ = open-loop transfer function
$G_1(s)$ = transfer function of existing plant
$G_2(s)$ = transfer function of series compensating networks
$K(s)$ = closed loop transfer function
q_i $(i=1,2\ldots n)$ = poles of closed-loop transfer function, $K(s)$
w_i $(i=1,2\ldots m)$ = zeros of closed-loop transfer function, $K(s)$
z_i $(i=1,2\ldots \mu)$ = zeros of plant transfer function, $G_1(s)$
z_i $(i=\mu+1,\mu+2,\ldots m)$ = zeros of compensator transfer function, $G_2(s)$
p_i $(i=1,2\ldots r)$ = poles of the plant transfer function, $G_1(s)$
p_i $(i=r+1,r+2,\ldots n)$ = poles of the compensator transfer function, $G_2(s)$
K_e = closed-loop gain constant
K_0 = open-loop gain constant

Paper 56-981, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted March 26, 1956; made available for printing August 31, 1956.

JOHN ZABORSZKY is with Washington University, and consultant to the McDonnell Aircraft Corporation, St. Louis, Mo.

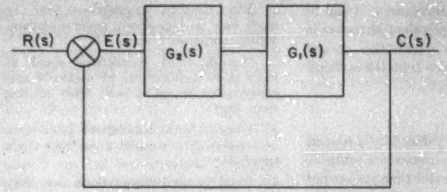


Fig. 1. Basic block diagram of a closed loop automatic control system with unit feedback

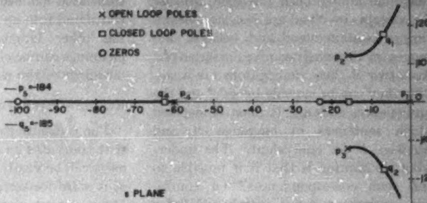


Fig. 2. Second root locus for illustrative example

K_d = displacement error constant
K_v = velocity error constant
K_a = acceleration error constant
T_p = time to reach first transient peak with step displacement input, Fig. 2
M = maximum transient overshoot with step displacement input
T_s = settling time of transient, to reach 2 per cent of initial displacement error
N = number of predominant oscillations in T_s
t = time
$s=\sigma+j\omega$ = complex frequency

$$\zeta=\frac{\sigma}{\sqrt{\sigma^2+\omega^2}},\quad \rho^2=\sigma^2+\omega^2$$

Discussion of Synthesis Techniques

Present synthesis techniques take one of two approaches: The historically older approach, as represented by the frequency response methods[1,2] of Nyquist, Bode, and Nichols, and by the root locus solution of Evans,[3] commences with the transfer function of the plant, and proceeds to estimate the corresponding closed-loop characteristics, the two being linked, (Fig. 1), by the relation

$$K(s)=\frac{G(s)}{1+G(s)} \tag{1}$$

If the closed loop characteristics fall short of the specifications, the open loop is supplemented by cascaded lead and lag compensating networks until satisfactory closed-loop performance is reached. However, the compensating networks must be preselected (for instance, lead or lag networks), rather than synthesized for the best achievement of the set goal in the closed loop. The objectives of the design are improvements in the closed-loop performance, yet the changes are made in the open-loop transfer function with no direct means, in most cases, to calculate what change to make to attain the goal.

The other approach[4] is the one proposed by Guillemin. This commences with the selection of poles and zeros of the closed-loop transfer function to conform to the specifications of steady, and transient, performance; bandwidth; etc. Once the closed-loop transfer function is chosen through its poles and zeros, the open-loop transfer function is determined by the inversion of equation 1 as

$$G(s)=\frac{K(s)}{1-K(s)} \tag{2}$$

This transfer function, then, is synthesized for the open loop with conventional network synthesis techniques.

This method in its original form can be used to determine open-loop transfer functions with real roots only, except through factoring equations of higher order. A more serious consideration is that it takes no advance cognizance of the plant and so yields and open-loop transfer function, the poles and zeros of which usually do not coincide with poles and zeros of the plant transfer function. The latter either must be cancelled by the compensating network of the open loop, or else the closed loop must be modified by shifting poles and zeros in it to cause poles and zeros of $H(s)$ in the open loop to coincide with those of the plant. The fact that the poles and zeros of the closed loop are less rigidly fixed than those of the plant probably accounts for the wider use of the methods in the first group.

General Synthesis Considerations

As may be observed, both of these approaches are somewhat one-sided; one is from the point of view of the open loop and the other from that of the closed loop. Yet the problem in its original form is highly symmetrical:

1. A set of specifications for the closed loop resulting in a desired transfer function $K(s)$ is given.
2. A set of components in the open loop combined into a plant with transfer function $G_1(s)$ is given.
3. A control network, or its transfer function $G_2(s)$, to match the foregoing two, is sought.
4. Open and closed loops are linked by the equation

$$K(s)+K(s)G_1(s)G_2(s)-G_1(s)G_2(s)=0 \tag{3}$$

Indeed, one of the most striking aspects of this statement of the problem is its symmetry, and it appears that the synthesis should not only take cognizance of this symmetry but should actually utilize it. A process to accomplish this will be discussed later in the paper.

Formalized Synthesis

It should be mentioned first that once the transfer functions $K(s)$ and $G_1(s)$ are fixed, the problem has only one solution for control network and it appears in a closed mathematical form as the solution of equation 3

$$G_2(s)=\frac{1}{G_1(s)}\,\frac{K(s)}{1-K(s)} \tag{4}$$

A corresponding network could be synthesized by conventional synthesis techniques.

A system so designed and, in general, only a system so designed would have exactly the closed-loop transfer function which was selected on the basis of the specifications and would at the same time incorporate the existing plant in the open loop. However, a closer study of equation 4 reveals that this is accomplished in an undesirable fashion. The term $1/G_1(s)$ in equation 4 means that in general the compensating network cancels every pole and zero of the plant by superimposing poles of $G_2(s)$ on the zeros of $G_1(s)$ and zeros of $G_2(s)$ on the poles of $G_1(s)$. Then, according to the second term on the right of equation 4, $G_2(s)$ introduces those poles and zeros into the open loop which are needed to produce $K(s)$. In other words, the existing open loop is cancelled and replaced by one to produce the selected closed loop. This could hardly be called an ideal method of synthesis.

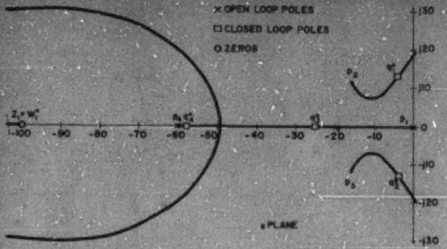


Fig. 3 (left). First root locus for illustrative example

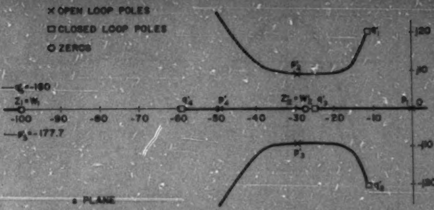


Fig. 4 (above). Reversed root locus for illustrative example

Integrated Synthesis

The breakdown of the formalized solution as a practical answer demands a more flexible interpretation of transfer functions $G_1(s)$ and $K(s)$. Actually, there may be some flexibility in the plant $G_1(s)$, and there certainly is substantial flexibility in the closed loop $K(s)$, since equally satisfactory closed-loop performance can be usually obtained by a number of various pole and zero configurations.

In this interpretation the closed loop and open loop are designed simultaneously to meet the specifications for both. The necessary flexibility is contributed by $K(s)$ in the closed loop and principally by $G_2(s)$ in the open loop, which still preserves the symmetry of the situation.

The solution itself can probably be best effected by an iteration process which fits the symmetry of the problem. The iteration can start either in the open or the closed loop. Suppose that the starting point is a $K(s)$ closed-loop transfer function which has been selected on the basis of the specifications.

1. The first step would be to prepare a reversed root locus according to Appendix II for the open-loop poles. If, now, the system is to be zero displacement error, a first-order pole in the open loop must be located at the origin. This pole fixes the closed-loop gain K_c and the latter locates all the other open-loop poles; see equation 26. Open-loop zeros coincide with closed-loop zeros, for unity feedback. See Fig. 1.

2. The open-loop poles and zeros thus obtained, of course, will not in general coincide with any of the poles and zeros of the plant, so the open-loop transfer function obtained must be modified. This will be done by shifting open-loop poles or zeros which are fairly close to poles and zeros of the plant to coincide with the latter. Where no pole or zero is close to those of the plant, poles and zeros are simply added in the open loop.

3. A conventional root locus for the closed-loop poles is now constructed on the basis of the revised open-loop poles and zeros, an open-loop gain is selected, and the poles of the closed loop are identified by conventional root locus techniques. Of course these poles will not now in general coincide with the starting set of step 1.

4. Consequently, modifications again must be made in the closed loop to meet specifications. First the specification values are evaluated from the results of step 3 and then poles and zeros are shifted according to need for meeting the specifications. The addition of further poles and zeros may be found necessary; this, of course, means a transition to a higher order $G_2(s)$ compensating network.

5. Open-loop poles and zeros are again obtained, as in step 1.

6. Poles and zeros of the plant are again moved back to where they belong, as in step 2.

7. Closed-loop poles and zeros again obtained, as in step 3.

8. Modifications in the closed loop made, as in step 4, etc.

Naturally the number of iterations actually required by a practiced designer will usually be small.

Clearly, the modifications in the closed loop must be selected and tailored to the individual case although there are some standard techniques[4,5,6] based on the equations of Appendix I.

As mentioned, the method is quite flexible as regards the starting point of the iteration. It can be started in the open loop rather than the closed with, for instance, the plant alone as the open-loop transfer function in the first step. Or it can be used to sharpen up a design obtained by some frequency response technique such as the Nyquist diagrams.

It should be noted that this proposed process still depends to a large extent on the experience and judgment of the designer. However, this is no deficiency, since it seems doubtful that a closed solution leading from the specifications of the two loops directly to the transfer function of a compensating network will ever be devised.

On the other hand, the proposed process does introduce symmetry and a higher degree of organization into the design, by substituting iteration for cut-and-try method. Looking at the problem from either the open or the closed loop the designer is forced to estimate what change to make in the loop selected for base in order to achieve a desired change in the other loop. But the 2-way use of root locus can answer the question as to what specific change should be made in either loop to effect a desired improvement in the other. Of course, there remains the compromise customary in engineering design as to how much of this change is permissible.

This presentation is on the basis of a unity feedback system; however, the extension to systems with networks in the feedback loop is obvious. Equally obvious is the adaptation of the use of other types of diagrams designed to locate roots.[7]

Illustrative Example

The advantages of the proposed method are apparent mainly in complex problems where the designer is likely to lose insight for the change required, for instance, in the open loop, in order to produce a desired improvement in the closed loop, or vice versa. In such cases the 2-way use of the root locus can supply the needed information. However, since complex problems would be unreasonably long and unwieldy for publication, only a simple illustrative example is presented here which could be solved with about equal ease by conventional techniques. Consequently, this example should be considered illustrative only in the sense of showing the steps in the process and not in the sense of demonstrating its usefulness.

Following are the specifications for the performance of a closed-loop automatic control system:

$K_a = \infty$
$K_v \geq 10$
$M \leq 25$ per cent
$T_p \leq 0.2$ second
$T_s \leq 1$ second
$\omega_{bw} \leq 30$ radians per second

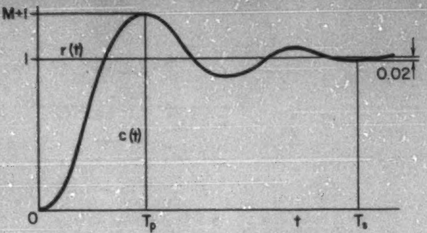


Fig. 5. General sketch of a transient response

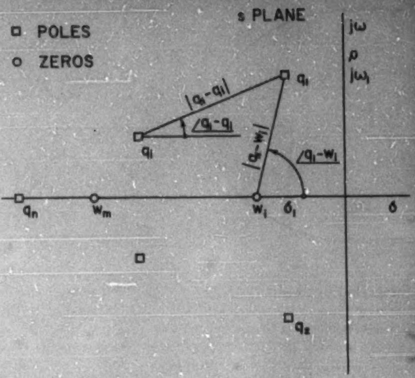


Fig. 6 (right). Closed-loop poles and zeros

The system must incorporate in its open loop a plant with the transfer function of

$$G_1(s)=\frac{s-z_1}{s(s-p_2)(s-p_3)(s-p_4)}$$

$$=\frac{s+100}{s(s+16-j12)(s+16+j12)(s+60)}$$

In this example the design is started by ascertaining the optimum performance possible without any compensating network. For this purpose a root locus based on poles p_1, p_2, p_3, and p_4, and zero z_1 of the plant itself, is constructed in Fig. 3. Now, an open loop gain of $K_0''=2{,}750$ is selected, which locates the closed loop poles at $q_1''=-4+j13$, $q_2=-4-j13$, $q_3''=-25$ and $q_4''=-57.7$. A closed-loop zero is present at $w_1''=z_1=-100$. The values involved in the specifications can now be evaluated by the equations of Appendix I; yielding

$K_d=\infty$
$K_v=6.25$
$M=31$
$T_p=29$
$T_s=1$ second
$\omega_{bw}=17$ radians per second

While these represent the optimum compromise which can be obtained from the plant alone simply by adjusting the gain, they do not satisfy the specifications. Consequently, the system will have to be modified by shifting the predominant pair of poles away from the origins and to the left into the position of $q_1'=-12.5+j20$ and $q_2'=-12.5-j20$. An additional zero will be required to make such shift possible; this will be placed to $w_2'=-27.5$. Finally, an additional pole is introduced at $q_5''=-180$ to maintain the excess of three in the number if poles over the number of zeros or, in other words, to avoid a compensating network with a pole at infinity. The other poles and zeros are left unchanged at $q_3'=q_3''=-25$, $q_4'=q_4''=-57.7$ and $w_1'=w_1''=z_1=-100$. The specification values for this set of closed-loop poles and zeros, as shown in Fig. 4, are

$K_d=\infty$
$K_v=16.8$ per second
$M=12$ per cent
$T_p=0.176$ second
$T_s=0.87$ second
$\omega_{bw}=26.5$ radians per second

These satisfy the specifications with some margin. Now, a revised root locus is constructed in Fig. 4, and the open-loop poles are established after the closed loop gain of $K_e=52{,}000$ is found from the condition of $p_1=0$, which is required by $K_d=\infty$. The resulting open-loop poles $p_2'=-29.9+j8$, $p_3'=-29.9-j8$, $p_4=-49$, and $p_5'=-177.7$, fail to coincide with the poles of the plant.

The next step is to move poles p_2', p_3', p_4' to coincide respectively with poles p_2, p_3, and p_4 of the plant, as shown in Fig. 2. To compensate for this move, zero w_2' is also moved to $w_2=z_2=-23$. The root locus constructed in Fig. 2 then yields with a gain of $K_0=2{,}700$, a new set of closed-loop poles at $q_1=-7+j17.5$, $q_2=-7-j17.5$, $q_3=-15.5$, $q_4=-62.5$, $q_5=-185$. These give

$K_d=\infty$
$K_v=13.1$ per second
$M=24.5$ per cent
$T_p=0.20$ second
$T_s=0.6$ second
$\omega_{bw}=23$ radians per second

These are within the specifications; consequently the iteration can now be discontinued. The compensating network is then represented by gain $K_0=2{,}700$, the pole $p_5=-184$, and the zero $z_2=w_2=-23$. Therefore

$$G_2(s)=2{,}700\,\frac{s+23}{s+184} \tag{5}$$

a simple lead network for this simple problem.

Conclusions

A viewpoint is presented, which gives more consideration than does the conventional viewpoint, to the double nature and the symmetry of closed-loop control system synthesis, with the controller as an element to match a set of specified components in the open loop to a set of performance specifications in the closed loop.

Also presented is a method of synthesis which assigns symmetrical roles to the specifications for open and closed loop in the design of a controller whether the method is used for design directly from the specifications or for refining a design for a controller which was obtained by conventional methods.

Appendix I. Selection of Closed-Loop Poles and Zeros

The basis of any selection of closed-loop poles and zeros must clearly be the relationships which exist between the pole and zero locations of the closed loop and such items of the performance specifications as bandwidth, steady-state errors or error constants, transient overshoot, and the speed and attenuation of the transient response. Such relationships have been developed and are summarized here. If the closed-loop transfer function is denoted, Fig. 1

$$K(s)=\frac{C(s)}{R(s)}=\frac{G_1(s)G_2(s)}{1+G_1(s)G_2(s)}$$

Vol. 7
Pt. 1
1956

$$= \frac{K_d}{1+K_d} - \frac{1}{K_v}s - \frac{1}{K_a}s^2 \ldots \quad (6)$$

or

$$K(s) = K_c \frac{(s-w_1)(s-w_2)\ldots(s-w_m)}{(s-q_1)(s-q_2)\ldots(s-q_n)} \quad (7)$$

Then, the steady-state errors, or rather the error constants, are determined as[4]

$$K_d = \frac{K_c \prod_{i=1}^{m} w_i}{(-1)^{n-m} \prod_{i=1}^{n} q_i - K_c \prod_{i=1}^{m} w_i} \quad (8)$$

$$\frac{1}{K_v} = \sum_{i=1}^{m} \frac{1}{w_i} - \sum_{i=1}^{n} \frac{1}{q_i}, \text{ if } K_d = \infty \quad (9)$$

$$\frac{1}{K_a} = \frac{1}{2}\left(-\frac{1}{K_v^2} \sum_{i=1}^{n} \frac{1}{q_i^2} + \sum_{i=1}^{m} \frac{1}{w_i^2} \right), \quad \text{if } K_d = \infty \quad (10)$$

The main characteristics of the transient response are determined,[5,8] Fig. 5

$$T_p = \frac{1}{\omega_0}\left(-\sum_{i=1}^{m} \underline{/q_1 - w_i} + \sum_{i=3}^{n} \underline{/q_1 - q_i} \right) \quad (11)$$

$$M = \prod_{i=1}^{m} \frac{|q_1 - w_i|}{|w_i|} \prod_{i=3}^{n} \frac{|q_i|}{|q_1 - q_i|} e^{-\sigma_1 T_p} \quad (12)$$

$$T_s = \frac{4}{\zeta \rho} \quad (13)$$

$$N = \frac{2}{\pi} \frac{\sqrt{1-\zeta^2}}{\zeta} \quad (14)$$

It is assumed for equations 11 through 14 that the closed-loop transfer function has a predominant pair of poles, i.e., two conjugate complex poles q_1 and q_2, Fig. 6, the real parts of which are much smaller (preferably at least one-third smaller) than the real part of the next pole; see Fig. 6. The assumption is justified because this type of arrangement is known to be the most desirable and also because it alone is amenable to reasonably simple evaluation on the basis of closed-loop poles and zeros. However, relations have been published for evaluating closed-loop response for systems without predominant poles.[8,9]

Bandwidth is defined as the frequency, ω_{bw}, where the closed-loop amplitude drops to $1/\sqrt{2}$ of its zero frequency value. It is the solution of the equation

$$1/\sqrt{2} = K_c \frac{\prod_{i=1}^{m} |j\omega_{bw} - w_i|}{\prod_{i=1}^{n} |j\omega_{bw} - q_i|} \quad (15)$$

or, approximately on the basis of the predominant roots only

$$\frac{\omega_{bw}^2}{\zeta^2} = (1-2\zeta^2) + \sqrt{2-4\zeta^2+4\zeta^4} \quad (16)$$

It must be remembered further that $G(s)$ must have an excess of poles over zeros, $(n-m)$, which is not smaller than the excess of poles over zeros in the given plant $G_1(s)$, unless the compensating network $G_2(s)$ is to have a pole at infinity.

The spirule[3] is a useful tool in evaluating equations 8 through 12 and solving equation 15. When the specifications for the closed-loop performance are given, closed-loop pole and zero locations can be selected by the aid of equations 8 through 15 to satisfy these specifications. This process has been discussed and illustrated[4,5,6] and will not be further discussed here.

Appendix II. Reversed Root Locus

According to equation 2, the only poles of the open-loop transfer function will be where

$$1 - K(s) = 0 \quad (17)$$

or

$$K(s) = 1 \quad (18)$$

This complex equation is equivalent to the following two real equations

$$\underline{/K(s)} = 0 \quad (19)$$

$$|K(s)| = 1 \quad (20)$$

or with equation 7

$$\sum_{i=1}^{m} \underline{/s - w_i} - \sum_{i=1}^{n} \underline{/s - q_i} = 0 \quad (21)$$

$$K_c \frac{\prod_{i=1}^{m} |s - w_1|}{\prod_{i=1}^{n} |s - q_1|} = 1 \quad (22)$$

Equation 19, or equation 21, of course, describes a root locus on which the sum of the angles is 0 degrees rather than 180 degrees. Like the conventional root locus, this too can be established by the spirule[3] or by phase-angle loci.[6]

Some helpful characteristics in establishing the reversed root locus are as follows:

1. The locus is symmetrical with respect to the real axis.

2. Branches of the locus terminate at closed-loop poles and zeros or at infinity.

3. The number of branches going to infinity equals $n-m$, the excess of closed-loop poles over closed-loop zeros if $m<n$. (23)

4. The branches going out to infinity will approach asymptotes which made the following angles with the real axis

$$k\frac{360}{n-m} \quad (24)$$

where $k = 0,1,2,\ldots(n-m-1)$ and $m<n$.

5. There is always at least one branch $(k=0)$ going to infinity. This branch is the section of the real axis lying to the right of the extreme right real pole or zero of $K(s)$.

6. All asymptotes meet at one point on the real axis, i.e., at the centroid

$$\frac{\sum_{1}^{n} Re|q_i| - \sum_{1}^{m} Re\,|w_i|}{n-m} \quad (25)$$

7. The initial angle of the locus at a complex pole is the sum of the angles of all vectors connecting all poles and zeros to the complex pole in question. Angles of vectors from zeros are taken positive and those from poles are taken negative.

8. The initial angle of the locus at a complex zero is equal to the sum of the angles of all vectors connecting all poles and zeros to the complex pole in question. Angles of vectors from zeros are taken negative and those from poles are taken positive.

9. At real single poles or zeros the locus starts along the real axis.

10. At real double poles or zeros the two branches of the locus start either along the real axis or perpendicular to it.

11. At very low closed-loop gain the open-loop poles almost coincide with the closed-loop poles, and with increasing gain they gradually shift to the closed-loop zeros or to infinity. The number of closed- and open-loop poles is the same, with unity feedback, Fig. 1, and $m<n$.

Since the preparation of this paper an extensive study[10] of the reversed root locus has been presented.

References

1. Principles of Servomechanisms (book), G. S. Brown, D. P. Campbell. John Wiley & Sons, Inc., New York, N. Y., 1948.

2. Servomechanisms and Regulating System Design (book), H. Chestnut, R. W. Mayer. *Ibid.*, 1951.

3. Control System Dynamics (book), W. R. Evans. McGraw-Hill Book Company, Inc., New York, N. Y., 1954.

4. Automatic Feedback Control System Synthesis (book), J. G. Truxal. *Ibid.*, 1955.

5. Synthesis of Closed Loop Systems Using Curvilinear Squares to Predict Root Location, D. W. Russel, C. H. Weaver. *AIEE Transactions*, vol. 71, pt. II, Jan. 1952, pp. 95–104.

6. Synthesis of Feedback Control System by Phase-Angle Loci, Yaohan Chu. *AIEE Transactions*, vol. 71, pt. II, Nov. 1952, pp. 330–39.

7. Automatic and Manual Control, A. Tustin, editor. *Symposium* articles by Kusters, Moore and Leonard. Academic Press, Inc., New York, N. Y., 1952.

8. The Effect of Pole and Zero Locations on the Transient Response of Linear Dynamic Systems, J. H. Mulligan, Jr. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 37, May 1949, pp. 516–29.

9. Further Effects of the Pole and Zero Locations on the Step Response of Fixed, Linear Systems, Armen H. Zemanian. *AIEE Transactions*, vol. 74, pt. II, Mar. 1955, pp. 52–55.

10. Root Locus in Feedback System Synthesis, John A. Aseltine. (Abstract), *Proceedings*, Institute of Radio Engineers, New York, N. Y., Mar 1956, p. 402.

Design of Wide-Band R-F Transformers Utilizing a Synthesized Equivalent Network

HITOSHI H. KAJIHARA
NONMEMBER AIEE

Synopsis: Design considerations of wide-band r-f (radio-frequency) transformers operating over a frequency range of 10 to 100 mc (megacycles) are reviewed. The complete basic equivalent circuit representation of transformers must be used in the analysis of the input impedance characteristics. A synthesized equivalent network is used in determining the critical transformer parameters. Experimental data and characteristics of a 25-watt, 20- to 70-megacycle unit are presented.

THE FLEXIBILITY of impedance transformation ratio, ease of fabrication, and physical compactness are inherent advantages of transformers. These factors favor its use over other methods of coupling antennae to transmitters. The knowledge of these transformer attributes and the recent development of low-loss ferrite core materials led to the investigation and realization of the 20- to 70-mc wide-band r-f transformer described in this paper. In conjunction with the development of this unit, a method of determining the critical transformer design parameters and of predicting its VSWR (voltage standing wave ratio)-bandwidth performance was formulated. It is the purpose of the paper to introduce this method of wide-band r-f transformer design, to review the design considerations of a specific r-f transformer, and to present experimental information useful in designs of other r-f units.

Discussion

A 25-watt, 20- to 70-mc r-f transformer was required to match a 92-ohm coaxial cable to a 1,200-ohm whip antenna. A maximum VSWR of 2, and a minimum efficiency of 80 per cent were required of the transformer. A satisfactory unit meeting these specifications was fabricated and field tested. Among the items of prime importance in obtaining a satisfactory unit were: 1. the input impedance characteristics to meet the VSWR requirements, 2. low core loss material, 3. the best winding arrangement for optimum coupling, and 4. the selection of conductors.

VSWR and Input Impedance Characteristics

Since the VSWR becomes of prime concern for satisfactory transmission of signals at these frequencies, the input impedance characteristic of the r-f transformer assumes primary importance. A typical input impedance characteristic of a terminated r-f transformer is shown in Table I. These impedances are given in terms of parallel resistive R_p and reactive X_p components. Examination of Table I shows that at the lower frequencies, the R_p is larger than the reflected load resistance R_L calculated on a turns-ratio-squared basis. The X_p is inductive. With increasing frequency, R_p decreases in magnitude. A frequency is reached where the input impedance becomes purely resistive. Above this frequency, the reactive component becomes capacitive. A second frequency is obtained where the input impedance is again purely resistive. Above the second resonant frequency, the reactive impedance becomes inductive. This input impedance characteristic can be duplicated by an equivalent network consisting of seven elements.

The schematic diagram of these transformers is shown in Fig. 1. The input impedance in terms of parallel resistive and reactive components is given by equations 1 and 2

$$R_p=\frac{A^4+\left\{B\left[C^2+\left(\frac{C}{D}-1\right)^2\right]-A^2\left[C+\frac{1}{D}\left(\frac{C}{D}-1\right)\right]\right\}^2}{A^2\left\{C^2+\left(\frac{C}{D}-1\right)^2\right\}}R_L \qquad (1)$$

$$X_p=\frac{X_{c1}X_p'}{X_{c1}-X_p'} \qquad (2)$$

where

$$X_p'=\frac{A^4+\left\{B\left[C^2+\left(\frac{C}{D}-1\right)^2\right]-A^2\left[C+\frac{1}{D}\left(\frac{C}{D}-1\right)\right]\right\}^2}{B\left\{C^2+\left(\frac{C}{D}-1\right)^2\right\}\times\left\{B\left[C^2+\left(\frac{C}{D}-1\right)^2\right]-A^2\left[C+\frac{1}{D}\left(\frac{C}{D}-1\right)\right]\right\}}X_{L1}$$

and

$$A=\frac{X_M}{R_L}$$

$$B=\frac{X_{L_1}}{R_L}$$

$$C=\frac{X_{L_2}}{R_L}$$

$$D=\frac{X_{c_2}}{R_L}$$

Synthesis of Equivalent Network

The simplest network which can make the reflected load resistance appear larger than the value calculated on a turns-ratio-squared basis is a network shown in Fig. 2. The resistive component is given by equation 3

$$R_p=\left\{1+\left(\frac{X_s}{R_L}\right)^2\right\}R_L \qquad (3)$$

To make R_p even larger, a capacitance C_0 can be placed in parallel with X_s. In the transformer, C_0 represents the direct capacitance from secondary to primary. The network is shown in Fig. 3. The R_p is given by equation 4

$$R_p=\left\{1+\left(\frac{X_s'}{R_L}\right)^2\right\}R_L \qquad (4)$$

where

$$X_s'=\frac{X_{c_0}X_s}{X_{c_0}-X_s}$$

A network which exhibits a characteristic of reducing R_p with increasing frequency over a range of frequencies is shown in Fig. 4. Beyond some frequency R_p will again increase. The R_p is given by equation 5

$$R_p=\left\{\left(\frac{X_s'}{R_L}\right)^2+\left(\frac{X_s'}{X_{c_2}}-1\right)^2\right\}R_L \qquad (5)$$

Paper 56-970, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1-5, 1956. Manuscript submitted June 7, 1956; made available for printing August 27, 1956.

Hitoshi H. Kajihara is with the Signal Corps Engineering Laboratories, Fort Monmouth, N. J.

The author wishes to acknowledge the contributions made by V. J. Kublin in the form of helpful and pertinent comments, suggestions, and ideas on the subject.

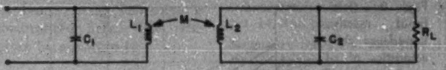


Fig. 1. Circuit representation of r-f transformers with coefficient of coupling less than unity

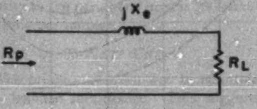


Fig. 2. Network for $R_p/R_L > 1.0$

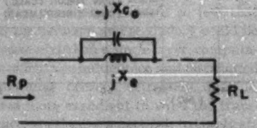


Fig. 3. Network for $R_p/R_L > 1.0$

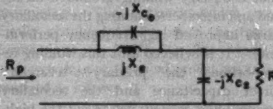


Fig. 4. Network for $R_p/R_L \leqslant 1.0$

To decrease R_p at the same rate as the actual r-f transformer, it was necessary to insert an equivalent core loss resistance in the network of Fig. 4.

The sign of the reactive component of the input impedance is determined by the primary self-inductance L_1, the primary distributed capacitance C_1, the leakage inductance and direct capacitance combination, and reflected secondary capacitance C_2. To duplicate the r-f transformer reactive impedance characteristics, an L_1 and C_1 is inserted in Fig. 4. Thus, the equivalent network which closely duplicates the r-f transformer impedance characteristics is shown in Fig. 5. It is recognized that this network is the complete basic equivalent network representation of transformers upon which simplifications are made for design of audio transformers. In r-f transformers whose coefficient of coupling is less than unity, similar simplifications cannot be made. Since the major portion of the elements in the equivalent network representation are parallel elements, it is convenient for computation purposes to convert the series leakage inductance and direct capacitance to equivalent parallel elements. If this conversion is performed, the circuit of Fig. 6 is obtained.

The leakage inductive impedance, reflected secondary capacitive impedance, and reflected load resistance are modified by K_0, K_1, and K_2, respectively.

$$K_0=\frac{\left(\frac{X_e'}{R_L}\right)^2+\left(\frac{X_e'}{X_{c_2}}-1\right)^2}{\left(\frac{X_e'}{R_L}\right)^2\left\{1+\left(\frac{R_L}{X_{c_2}}\right)^2\right\}} \tag{6}$$

$$K_1=\left(\frac{X_e'}{R_L}\right)^2+\left(\frac{X_e'}{X_{c_2}}-1\right)^2 \tag{7}$$

$$K_2=K_1 \tag{8}$$

To calculate the numerical values of R_p and X_p of Fig. 6, it is simplest to use the input admittance equation. This is given by equation 9

$$Y_p=\frac{1}{R_p}-j\frac{1}{X_p}=\left(\frac{1}{K_2R_L}+\frac{1}{R_{c_1}}\right)-j\left(\frac{1}{X_{L_1}}-\frac{1}{X_{c_1}}+\frac{1}{K_0X_e'}-\frac{1}{K_1X_{c_2}}\right) \tag{9}$$

The substitution of typical r-f transformer parameter values in equation 9 shows that the first resonant frequency is determined primarily by the primary self-inductance and the distributed capacitances. This is a parallel resonance in Fig. 5, and the transformer performance is not affected. The second resonant frequency is determined primarily by the leakage inductance and direct capacitance. This is a series resonance in Fig. 5. It is detrimental to transformer performance from the power transfer consideration. Consequently, for wide-band operation of transformers it is very important to minimize these two parameters.

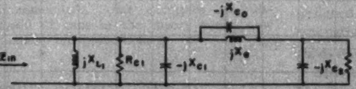


Fig. 5 (left). Complete equivalent network of r-f transformers

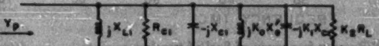


Fig. 6 (left). Equivalent network of r-f transformers used for calculations of critical parameters

Table I. Typical Input Impedance Characteristics of Terminated R-F Transformer

Load Reflected into Primary, 78 Ohms on Turns-Ratio-Squared Basis

f, Megacycles	R_p, Ohms	X_p, Ohms
15	87	$j133$
20	84	$j214$
30	76	$j670$
38	67	∞
40	65	$-j4980$
50	54	$-j3520$
53	52	∞
60	47	$j1160$
70	42	$j212$
85	43	$j74$
90	47	$j63$
100	64	$j54$

CONSTRUCTION OF EQUIVALENT NETWORK

The R_p and X_p measurements of a terminated r-f transformer are made over the appropriate frequency range. If at the lower frequencies the R_p is larger than R_L, leakage inductance is present. The leakage inductance may be measured by one of two methods. The first method is the conventional bridge measurement technique with the secondary winding short-circuited. A Q meter may also be used for the measuring instrument. A second method is to measure the self-inductances of the windings, and the series-aiding and series-opposing inductances of the two windings on an RX (resistance-reactance) meter or a Q meter. From these measurements, the mutual inductance is calculated and the leakage inductance determined.

In all the inductance measurements at high frequencies, corrections may be neces-

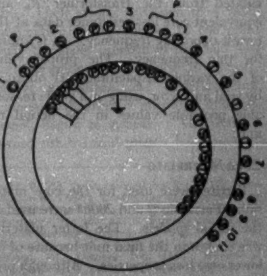


Fig. 7. Winding arrangement of final r-f transformer

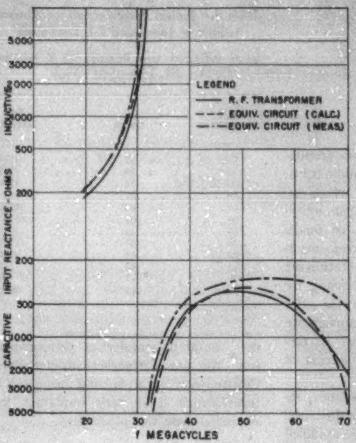
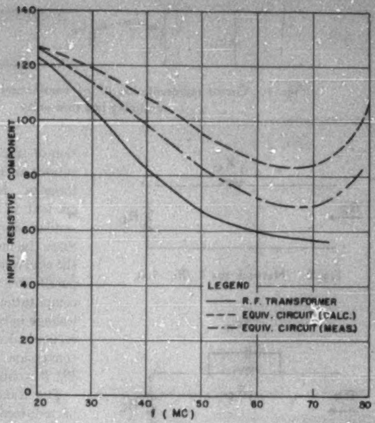


Fig. 8 (left). Input reactive impedance characteristics of r-f transformer and synthesized network

Fig. 9 (right). Input resistive characteristics of r-f transformer and synthesized network



sary because of distributed capacitance effects. The measurements should be made at several frequencies, a minimum of two. The true inductances and distributed capacitances can be calculated with equations 10 and 11. The subscripts 1 and 2 refer to two frequencies f, where f_2 is greater than f_1.

$$L_{\text{act}} = L_{\text{meas2}} \frac{1-\left(\frac{f_2}{f_1}\right)^2}{1-\frac{L_{\text{meas2}}}{L_{\text{meas1}}}\left(\frac{f_2}{f_1}\right)^2} \tag{10}$$

$$C_d = \frac{L_{\text{meas1}} - L_{\text{act}}}{L_{\text{meas1}} \times L_{\text{act}}(2\pi f_1)^2} \tag{11}$$

For the purpose of obtaining an approximate leakage inductance for the equivalent network, it is simpler and quicker to use equation 3, and a value of R_L calculated on a turns-ratio-squared basis. The value of the reflected load resistance for use in the equivalent network is calculated from the transformer R_p and equivalent core loss resistance data at the lowest operating frequency. L_1 is known for this frequency. The proper value of the various capacitances is obtained by trial and error, using as a guide reasonable probable values in the actual r-f unit.

Core Materials

Ferrites were used for the core material. Ferramic *Q* and *2006** were used in the initial models. The latter material was chosen in the final unit because of its lower core loss properties. Although precise core loss measurement was not attempted at actual operating flux condition, the core temperature rise above ambient was noted. The relative temperature rise of *2006* to Ferramic *Q* at 70 megacycles and 10 gausses was about one to five. The core temperature rise of *2006* was 15 degrees centigrade.

Winding Arrangement

Various winding arrangements were investigated to obtain maximum coupling. A schematic diagram of the winding arrangement used in the final model is shown in Fig. 7. This arrangement yielded best performance. It was found that a tertiary winding grounded at one end and interspersed among the secondary turns improved high-frequency performance. It is believed that this winding reproportions the primary-to-secondary direct capacitance and the secondary interwinding capacitance to aid the high-frequency operation.

Selection of Conductors

The basic equation for the depth of penetration of current d_p into conductors was used as a guide in the selection of conductors. This equation neglects prox-

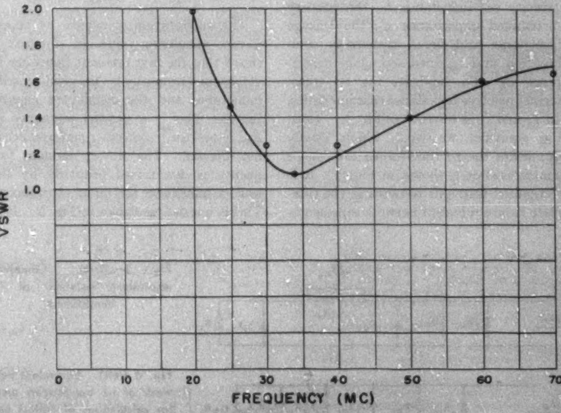


Fig. 10. VSWR versus frequency of final r-f transformer

---

* Designation used by General Ceramics Corporation, Keasbey, N. J.

Vol. 7
Pt. 1
1956

imity effects but extreme exactness is perhaps unwarranted here. The relationship is given by equation 12.

$$d_p = \frac{2.61}{\sqrt{f\text{mc}}} \text{ mils} \tag{12}$$

With the use of equation 12, a wire diameter is calculated so that the current density in the peripheral area is 200 circular mils per ampere.

DESIGN OF 25-WATT UNIT

The primary concern was to maintain low leakage inductance. This suggests keeping the turns to a minimum and investigating the optimum winding arrangement. Since the total turns required from the flux consideration were so few that both the primary and secondary could be accommodated on a single layer, computation of leakage inductance and distributed capacitance became extremely difficult. It was more expedient to wind the transformer, synthesize a network to duplicate its impedance characteristics, and then study the latter to determine, and make readjustment of, the critical transformer parameters. The network which was finally obtained to duplicate the transformer is shown in Fig. 5. The winding and core data of the final model are summarized as follows.

Core material: *2006*.
Core size: $1^{15}/_{16}$-inch outside diameter, $1^{7}/_{16}$-inch inside diameter, 3/8-inch height.
Number of cores: 2 per unit.
Primary turns and wire size: 3 turns of No. 14, 3 in parallel.
Secondary turns and wire size: 11 turns of No. 14, single.

EXPERIMENTAL DATA

The values of the components used to simulate the actual transformer are as follows, in μh (microhenrys) and $\mu\mu f$ (micromicrofarads):

$L_1 = 1.0\ \mu h$
$L_e = 0.15\ \mu h$
$C_0 = 10.0\ \mu\mu f$
$C_1 = 20.0\ \mu\mu f$
$C_2 = 15.0\ \mu\mu f$
$R_L = 134.0$ ohms

The input impedance characteristics of the actual unit and the measured and calculated impedance characteristics of the equivalent network are shown in Figs. 8 and 9. The R_p measurements on the equivalent network were taken without incorporating an equivalent core-loss resistive component since a different value is required at each frequency. This accounts for the discrepancy between the R_p of the equivalent network and the r-f unit. The VSWR-versus-frequency characteristic of the r-f transformer is shown in Fig. 10.

MEASUREMENT TECHNIQUES

An *RX* meter was used to measure the R_p and X_p. The VSWR was calculated from the Smith chart and also measured under actual operating conditions. Measurements at rated power were obtained by using Signal Corps *AN/TRC* transmitters. A specially constructed copper tubing with a center conductor was used to simulate a 92-ohm coaxial cable. The length was over a quarter wave length at the lowest frequency. Holes were perforated every 6 inches along the tube to permit entrance of a voltmeter probe for VSWR measurements.

Conclusions

The complete basic equivalent circuit representation of transformers must be used in the design and analysis of r-f transformers whose coefficient of coupling is less than unity. The critical design parameters may be established either by calculations or measurement on this network.

To obtain wide-band operation of r-f transformers in the 10- to 100-mc region the primary consideration is to minimize the leakage inductance and the primary-to-secondary direct capacitance.

References

1. ALTERNATING CURRENT CIRCUIT, Russell M. Kerchner, George F. Corcoran. John Wiley & Sons, Inc., New York, N. Y., second edition, 1943, pp. 108–109, 201–10.

2. RADIO ENGINEERING, Frederick E. Terman. McGraw-Hill Book Company, Inc., New York, N. Y., third edition, 1947, p. 21.

The Communication Engineer's Needs in Information Theory

W. T. REA
MEMBER AIEE

Synopsis: Closer rapport between the mathematician and the communications engineer would permit information theory to be exploited more effectively. Improvement in the number and type of educational courses, symposia, and publications is recommended. Some examples are discussed of gaps in the theory which now handicap its application.

THE COMMUNICATION engineer should rely heavily upon information theory in planning new communication systems. Before its advent, much of what he did was likely to be intuitive. He knew, or sometimes "sensed," that certain choices led to superior results, but he had no satisfactory standard against which to measure the performance of the systems he designed nor an adequate method of comparing them, with any degree of rigor, against other possible systems.

Although Nyquist[1,2] and Hartley[3] laid the foundations of information theory during the mid-1920's, too few communications engineers became familiar with their work and its implications to exert a profound impact upon communications technology. But the comprehensive formulation which this theory received in the work of Shannon[4] and Wiener[5] in 1948 moved the art of communications a long step toward the status of a science. Since that time, a considerable number of mathematicians possessing great power and insight have been operating in this field. Their efforts have led to a rapid ramification of the subject to the point where the sheer breadth of the field now tends to present an obstacle to comprehensive exposition and understanding. Some specialization is indicated and has, indeed, begun to appear. The trend is toward subdivision[6] into:

Paper 56-653, recommended by the AIEE Communication Theory Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956, and re-presented at the Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted April 16, 1956; made available for printing July 26, 1956.

W. T. REA is with Bell Telephone Laboratories, Inc., Murray Hill, N. J.

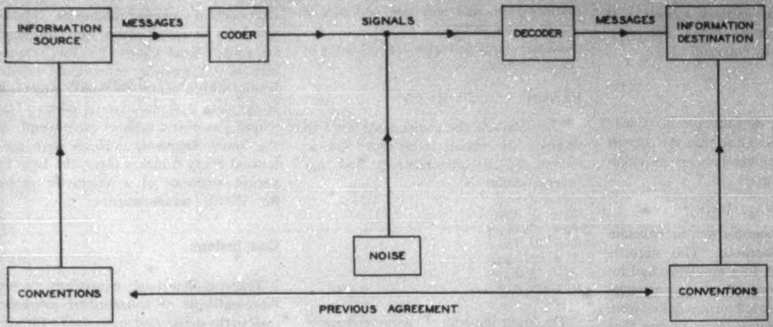

INFORMATION SOURCE
MESSAGES
CODER
SIGNALS
DECODER
MESSAGES
INFORMATION DESTINATION
NOISE
CONVENTIONS
PREVIOUS AGREEMENT
CONVENTIONS

Fig. 1. One-way communication

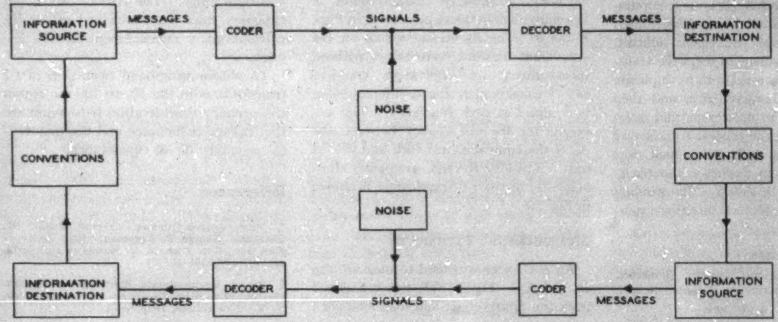

INFORMATION SOURCE
MESSAGES
CODER
SIGNALS
DECODER
MESSAGES
INFORMATION DESTINATION
NOISE
CONVENTIONS
CONVENTIONS
NOISE
INFORMATION DESTINATION
MESSAGES
DECODER
SIGNALS
CODER
MESSAGES
INFORMATION SOURCE

Fig. 2. Two-way communication

1. Basic theory, for the mathematician interested primarily in the further development and extension of the subject.

2. Coding theory, of chief interest to the coding expert and the cryptographer.

3. Communication (modulation) theory, for those engaged in the synthesis and analysis of communication systems.

4. Biophysical theory, for the biophysicist, the experimental psychologist, the neurophysiologist, and the specialist in human engineering.

5. Linguistic theory, which appeals to the semanticist, the comparative philologist, and the researcher into the origins of language.

Indeed, the proceedings of some information theory symposia seem to encompass the areas of feedback and computers.

The second and third areas just listed are of principal interest to the communications engineer, and he has benefited greatly from the principles established. However, when he tries to use the mathematician's work as a guide in synthesizing, developing, and designing new systems, he is likely to find it incompletely applicable to his problems, and to wish that he either knew more or had access to an expert consultant.

With so many aspects of modern life dependent upon efficient communication, it seems important to do whatever may be necessary to improve the applicability and increase the utilization of information theory in the communications art. One need that has become evident is for more and better courses in Information Theory. Although a scattering of courses in the subject has been instituted at the graduate level, many more are needed here, as well as some at the undergraduate level if engineering graduates in the communications option are to be equipped with the basic techniques they so vitally need.

More ready access to new results in information theory would be helpful. Much has already been accomplished in this direction by annual symposia, the proceedings of which have been published and, in some cases, distributed to members of technical groups. However, more and better papers are needed for delivery at technical meetings and publication in technical magazines. A particularly evident gap exists in the literature; only one or two comprehensive books on the subject have been published.

A related need is for more thorough exposition. The new results attained by the mathematician's work in the field must be developed and processed to elucidate their application to specific situations. Interpreters of the experts must bring the important principles of information theory within the engineer's frame of reference by expressing them in terminology meaningful to him. The more specialized and abstruse the new work becomes, the more such exposition is needed. It is important that publication of tutorial papers should follow, without too great a lag, the appearance of significant original results.

For maximum benefit to the user, and through him to the world at large, the subject must be developed in directions useful in the communications art. Light must be thrown into the theory's dusty corners. Gaps must be filled that now stand as obstacles in the way of the user.

Some examples are discussed in the remainder of this paper.

Two-Way Communication

In dealing with practical communication situations, it is often difficult to relate their mode of action with that of the mathematical models upon which information theory is based.

One of these models is depicted in Fig. 1. It postulates an information source that emits, presumably in response to some series of stimuli, a flow of messages. At some destination the messages are interpreted in accordance with a set of conventions upon which the source and the destination have previously agreed. The quantity of information each message contains is defined as the negative of the logarithm of the probability of the particular message being emitted. Presumably the less probable messages will be more unexpected to the destination, and will thus impart more information. But it is worth noting that the unexpectedness of any message will, in some sense, depend upon the comprehensiveness of the conventions agreed upon.

The channel that transfers the messages from source to destination comprises an encoder which translates messages into signals, a noise source interfering with the signals, and a decoder translating the sum of signal and noise into messages. The capacity of the channel is affected by the conditional probability of the same message being received as that which was sent.

In actuality, the most common communication process functions in the manner illustrated by Fig. 2. This involves an interchange of messages between two individuals, each of whom includes an information destination coupled with an information source. During the 2-way communication process, which has strong feedback coupling, the "conventions" are continually changing, and the messages emitted are strongly influenced by the information content of previously received messages. This 2-way flow acts both to correct transmission errors and to remove imperfections from the structure of conventions.

If it is assumed that such a process has the purpose of maximizing some benefit, such as the knowledge of one or both of the participants, or enlarging their area of agreement, its theoretical treatment might involve a number of disciplines, such as:

1. Decision theory, in the coupling from destination to source.
2. Game theory, for, conceivably, the joint action to maximize mutual gain might be treated as a non-zero-sum game.
3. Feedback theory, which would be involved by the dynamics of the process.
4. Communication theory, needed for the efficient employment of the communication channel.

The step from the unidirectional to the bidirectional model is a large one and it may prove difficult. However, it is much needed and it offers a challenging opportunity for mathematical research. It is not impossible that the regrettable trend toward 1-way communication systems in the military, commercial, and administrative fields may be influenced by the concentration of theory on unidirectional principles and the lack of a bidirectional theory. Certainly this trend evidences insufficient appreciation of the importance of 2-way communication in reducing error and promoting understanding.

Once success has been attained in working out a comprehensive theory applicable to bidirectional systems, extension to multiway systems (which have interesting properties of their own) will very likely be found relatively easy.

Role of Delay

One of the facets of information theory most meriting investigation bears on the role of delay. A principle laid down by Shannon states that as the efficiency of communication is increased by removing redundancy, more delay is incurred in the encoding and decoding processes. This is somewhat analogous to the delay that accumulates as the transmission band of a wave filter is narrowed.

In actual practice, the usefulness of information often depends upon its freshness. Because delay handicaps 2-way communication, it would be helpful to the communications engineer if the theory provided a quantitative measure of the amount of delay that must be incurred in order to obtain a given improvement in the efficiency of communication. A promising start in this direction has recently been made by Shannon.[7,8]

Meaning and Utility of Information

The concept of quantity of information, as based on statistical or probabilistic grounds, is usually adequate in dealing with transmission systems. However, when the communications engineer becomes concerned with the larger systems that gather, transmit, process, and display information, he may find himself faced with problems that involve the meaning or the value of information. For example, he must avoid choices that deteriorate meaning or value. Delay, as mentioned, may be an important parameter affecting the value of information.

Until comparatively recently, the theory has tended to ignore such matters. Possibly some principles useful in dealing with the meaning or utility of information might be uncovered in an attack on the problem of 2-way communication. For example, since the utility of information depends upon its relevance, high value may require an optimum degree of cross-correlation with information previously received or stored. Too low a degree would intimate irrelevance, whereas too high a degree would indicate excessive redundancy. Recent work[9] by MacKay relates "meaning" to the change in the recipient's "conditional probability matrix."

Digital Codes for Reducing Magnitude of Error

In sending quantitative information by pulse code modulation, the instantaneous magnitude of a parameter is usually encoded as an n-digit binary number. A binary signal representing this number is then modulated onto some carrier wave for transmission.

A transmission error in the signal element associated with the most significant digit of the binary number will cause an error of half the maximum magnitude, whereas an error affecting the least significant digit will produce an error of only $1/2^n$ the maximum. Thus the signal element representing the most significant digit bears 2^{n-1} as much burden as that representing the least significant digit.

It does not seem proper that elements bearing different burdens should be equally susceptible to error. In a well-designed transmission system, it might be preferable to have each portion of the signal protected against error to an extent roughly proportional to the burden it bears, so that failure would be more likely to occur simultaneously over the whole transmission system. The discrepancy between burden and protection might be alleviated by distributing the burden more equitably or by strengthening the portions carrying heavier burdens.

For example, to distribute the burden more equitably, the permutations available in the code might be associated with the various possible magnitudes of the parameter in such a way as to reduce the consequences of the more probable sorts of error. Some examples of codes of this sort have been worked out by Kautz,[10] who cites them as having a low "confusion rating" because the more probable sorts of error will do comparatively little damage to the message. A set of general rules is needed for deriving codes of this sort, applicable to linear and circular scales of various degrees of precision. Also, much theoretical work is needed in evaluating the transmission improvement attainable by exploiting the principle of proportionality between protection and burden.

Abnormally Distributed Noise

Communications engineers often feel the need for methods of dealing with the effects of noise having a statistical distribution of amplitudes that is neither Gaussian nor Rayleigh. Many actual communication media are subject to "spiky" or "impulse" noise. Theoretical study of such cases would be greatly facilitated by the development of a readily manipulable analytic function to represent an amplitude distribution characteristic of this sort of noise.

Conclusions

To bridge the gap between those who forge the tools of information theory and those who use them, the communications engineer must understand the mathematician's results and the mathematician must be apprised of the communications engineer's needs. As intermediaries, more applied mathematicians are needed who are versed in communications engineering and more communications engineers who are educated in the necessary mathematics. They are essential to serve as the channel of what should be an efficient 2-way communication system.

References

1. CERTAIN FACTORS AFFECTING TELEGRAPH SPEED, **H. Nyquist.** *Bell System Technical Journal*, New York, N. Y., vol. 3, 1924, pp. 324–46.

2. CERTAIN TOPICS IN TELEGRAPH TRANSMISSION THEORY, **H. Nyquist.** *AIEE Transactions*, vol. 47, Apr. 1928, pp. 617–44.

3. TRANSMISSION OF INFORMATION, **R. V. L. Hartley.** *Bell System Technical Journal*, vol. 7, 1928, pp. 535–63.

4. A MATHEMATICAL THEORY OF COMMUNICATION, **C. E. Shannon.** *Ibid.*, vol. 27, 1948, pp. 379–423, pp. 623–56.

5. THE EXTRAPOLATION, INTERPOLATION AND SMOOTHING OF STATIONARY TIME SERIES, WITH ENGINEERING APPLICATIONS (book), **N. Wiener.** John Wiley & Sons, Inc., New York, N. Y., 1949.

6. A BIBLIOGRAPHY OF INFORMATION THEORY—COMMUNICATION THEORY—CYBERNETICS, **F. L. Stumpers.** *Transactions*, Professional Group on Information Theory, Institute of Radio Engineers, New York, N. Y., Nov. 1953.

7. THE RATE OF APPROACH TO IDEAL CODING, **C. E. Shannon.** *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1955, p. 47.

8. SOME RESULTS IN CODING THEORY, **C. E. Shannon.** *Ibid.*, 1956, p. 126.

9. SYMPOSIUM ON INFORMATION THEORY, **Colin Cherry,** editor. "The Place of Meaning in the Theory of Information," **D. M. MacKay.** Academic Press, New York, N. Y., 1955, pp. 215–25.

10. OPTIMIZED DATA ENCODING FOR DIGITAL COMPUTERS, **W. H. Kautz.** *Convention Record*, Institute of Radio Engineers, New York, N. Y., pt. 4, 1954, pp. 47–57.

Analysis of Linear Sampled-Data Systems with Finite Pulse Width: Open Loop

G. FARMANFARMA
STUDENT MEMBER AIEE

Synopsis: An exact method for analysis of sampled-data systems with finite pulse width is presented. The results make it possible to obtain the output of such systems in a closed form, as a continuous function of time, and without recourse to any approximations. (The closed form is that which does not involve summation of series.) The analysis is based on introducing a new transform method, analogous to the z-transform, except that the actual pulse width and pulse shape are considered. This is called the p-transform where p denotes pulse width. Furthermore, by utilizing the modified z-transform, a method for obtaining the inverse Laplace transform of functions involving both ϵ^{Ts} and s is developed. This technique has been advantageously applied to evaluate the inverse p-transform in a closed form. (The term "modified z-transform" was introduced in reference 1.) The limiting cases of the p-transform are investigated. It is concluded that the p-transformation method provides a powerful technique for analysis of sampled-data systems with finite pulse width.

ANALYSIS and design of sampled-data control system have been the subject of considerable investigation in the last few years. The methods of analysis are based on: 1. Laplace transform and the use of impulse response,[2] 2. Laplace transform leading to the frequency response approach,[3] and 3. z-transform[4–6] and the modified z-transform.[1,7]

The impulse response is a logical application of the superposition principles. It is based on adding the effect of each successive pulse, either graphically or numerically. Basically it is a tool for analysis. It can give the continuous output and has the advantage that it is applicable to both the periodic and nonperiodic sampled-data systems. However, its scope of application is limited because of the labor involved.

The second method is based on the infinite series equivalent of the z-transform.[5,6] Theoretically, it can yield the continuous response. But in general the output is in the form of an infinite series, and hence it must be approximated. The degree of accuracy depends on the number of terms included and the rapidity with which the infinite series converges. The computational labor renders this method inferior to both the impulse response and the z-transform technique.

Development of the z-transform in recent years has made this a powerful tool for analysis. The z-transform has one serious shortcoming: it can yield the output only at the sampling instants $t=nT$. However, at present there are several methods for determining the response between sampling instants utilizing z-transform.[1,7–10] Among these the modified z-transform[1,7] is the most valuable.

The modified z-transform is an important tool in the analysis and synthesis of sampled-data systems. Its application to analysis makes it possible to obtain the output as a continuous function of time, leading to determination of important quantities such as the peak time, peak overshoot,[1] etc. Its recent extension to synthesis[11] has shown it to be useful in system design, enabling the designer to meet rigid specifications. The modified z-transform as a mathematical tool has

---

Paper 56-982, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted May 30, 1956; made available for printing August 30, 1956.

G. FARMANFARMA attends the University of California, Berkeley, Calif.

The author is indebted to Dr. E. I. Jury for his help and encouragement throughout the preparation of this paper. This research was supported by the United States Air Force, through the Air Force Office of Scientific Research, Air Research and Development Command.

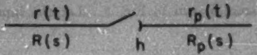


Fig. 1 (left). Ideal sampler with pulse width h

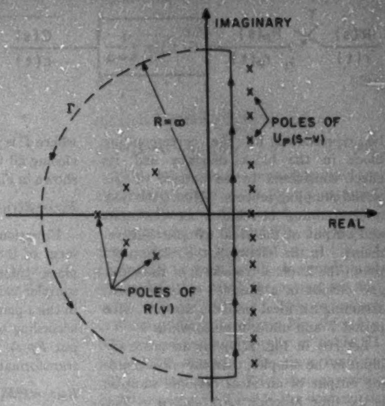


Fig. 3. Path of integration Γ in left half plane of complex variable v

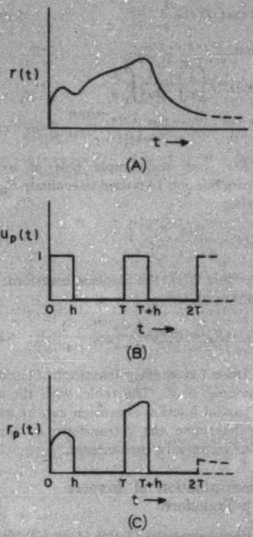


Fig. 2. A—Input signal to sampler. B—Unit pulse function. C—Sampler output constructed by multiplying input by unit pulse function

many applications outside the field of sampled-data control systems. Among these are its use in evaluating inverse Laplace transforms[13] (see the Appendix), and its application to design of continuous systems employing digital computers.[11]

The foregoing methods are all applicable to solution of sampled-data systems on the assumption of zero pulse width. Essentially the solutions are exact only if: 1. the sampler transforms the finite amplitude of input signal to true impulses of equivalent area (physically impossible), or 2. the sampler is followed by a hold circuit. (A hold circuit or boxcar generator is a circuit which maintains the output at the level of the sample until the new sample is taken.) Negligence to satisfy this fundamental requirement has led to some confusion in the past,[9] resulting in solutions which are known to be physically impossible.

It is known that with a finite pulse width, and in the absence of a hold circuit, the z-transform (or any of the foregoing methods) can be used to obtain an approximate solution.[4] This is based on the assumption that the pulse width is small compared to the time constants of the system. However, the degree of approximation has not been thoroughly investigated. Nor has there been a method for determining the response when the pulse width is comparable to the system time constants.

Investigating the effects of pulse width is of considerable importance in view of its engineering applications. For instance, a pulsed network in a servo loop might provide a desirable method of compensating the system. Also, techniques developed for the analysis of finite pulsed systems might eventually be applied to the analysis of relay servomechanism and other servo systems employing discontinuous elements such as biased magnetic amplifiers or vacuum-tube devices.

In this paper a new transform called the p-transform (p for pulse width) has been introduced for the exact analysis of finite pulsed linear systems. A general relation between the inverse Laplace transform and the inverse modified z-transform has been developed[13] (see the Appendix) and utilized to yield the inverse p-transform in a closed form. Tables of p-transform and inverse p-transform are given. With the aid of these tables it is possible to compute the output of a finite pulsed system as a continuous function of time. Numerical examples are worked to show the applications of these tables. The limiting cases of the p-transform have been investigated. It is shown that as the pulse width h approaches the sampling period T, the p-transform approaches the ordinary Laplace transform for the continuous system. This is an expected result and can serve as a valuable check in many problems. Furthermore, the extent of the error caused by the use of z-transform approximation and its validity are thoroughly investigated. Numerical example is worked to show the error that can be caused when the z-transform approximation is applied to systems having discontinuous impulse response.

The p-transform is a general tool for treatment of finite pulsed systems. It has been applied to solutions of feedback problems. However, much remains to be done before a complete report can be made. Hence, this paper is limited to the applications of the p-transform to solution of open-loop systems with a periodic sampler having uniform pulse width (i.e., all pulses are of the same width h).

Introduction to p-Transform

As it is often stated, the basic difference between continuous system and a finite pulsed sampled-data system is due to the presence of the sampler in the latter. Thus any general investigation of the pulsed system must start with the study of the sampler, and the relation between its input and output time functions. Fig. 1 illustrates an ideal sampler, ideal because it is assumed that it can make or break instantaneously. If the sampling equipment does not possess this ideal characteristic, then it must modify the signal according to some known function. It is clear that under these conditions the modifying function can

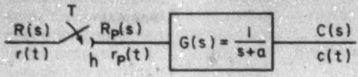


Fig. 4. Finite pulsed system

be represented by the corresponding block in the block diagram and its effect considered to be outside of the actual sampling process. Hence the task of determining the relation between input and output of the ideal sampler still remains. In the following, the relation between the Laplace transform of the input and output of a sampler is determined, assuming an ideal periodic sampler with period T and uniform pulse width h.

Let $r(t)$ in Fig. 2(A), be an arbitrary input to the sampler, then the corresponding output of an ideal periodic sampler is the time function $r_p(t)$, shown in Fig. 2(C). Inspection shows that it is possible to reconstruct $r_p(t)$ by multiplying the input $r(t)$ by a time function $u_p(t)$ as shown in Fig. 2(B) where

$$\begin{aligned} u_p(t)&=1 \qquad nT\leq t\leq nT+h \\ u_p(t)&=0 \qquad nT+h\leq t\leq (n+1)T \\ & \qquad\quad n=0,1,2,3\ldots \end{aligned} \tag{1}$$

Fig. 2 shows that when $r(t)$ is multiplied by the unit pulse function $u_p(t)$, the resultant is the time function $r_p(t)$ indicated in Fig. 2(C), which is identical to the sampler output. Thus, in effect, the sampler may be regarded as a multiplier in the time domain, multiplying its input by the unit pulse function $u_p(t)$. Hence, the relation between input and output in the time domain may be written as

$$r_p(t)=r(t)u_p(t) \tag{2}$$

where $r_p(t)$ is the sampler output. The Laplace transform of equation 2 is

$$\mathfrak{L}[r_p(t)]=\mathfrak{L}[r(t)u_p(t)] \tag{3}$$

Complex convolution[6,13] can be used to evaluate equation 2 as follows: Let

$$\mathfrak{L}[r_p(t)]=R_p(s) \tag{4}$$

$$\mathfrak{L}[r(t)]=R(s) \tag{5}$$

and

$$\mathfrak{L}[u_p(t)]=U_p(s)=\frac{1-\epsilon^{-hs}}{s(1-\epsilon^{-Ts})} \tag{6}$$

Then the Laplace transform of the output $R_p(s)$ may be expressed in terms of $R(s)$ and $U_p(s)$ as

$$R_p(s)=\frac{1}{2\pi j}\int_\Gamma R(v)U_p(s-v)dv \tag{7}$$

Substituting equation 6 in 7 results in

$$R_p(s)=\frac{1}{2\pi j}\int_\Gamma R(v)\frac{1-\epsilon^{-h(s-v)}}{(s-v)[1-\epsilon^{-T(s-v)}]}\,dv \tag{8}$$

where Γ is the contour of integration enclosing all the poles of $R(v)$ to its left,[6] shown in Fig. 3, and

$$R(v)=R(s)|_{s=v} \tag{9}$$

Equation 8 expresses the Laplace transform of the output in terms of the Laplace transform of input. Hence the sampler can be regarded as an operator in the s-plane transforming the input $R(s)$ according to equation 8 to give the output $Rp(s)$. For brevity this is called p-transformation and is denoted by

$$R_p(s)=\mathrm{P}[R(s)] \tag{10}$$

Its inverse is denoted by

$$r_p(t)=\mathrm{P}^{-1}[R_p(s)] \tag{11}$$

It is evident from the preceding equations that the p-transform is only a particular form of the Laplace transform. Thus all operations which are valid for Laplace transformations are identically valid for the p-transformation.

Equation 8 can be readily evaluated by the residue method. Assuming

$$R(s)=\frac{A(s)}{B(s)} \tag{12}$$

where $A(s)$ and $B(s)$ are polynomials in s, then $R_p(s)$, the Laplace transform of the sampler output, is given by

$$R_p(s)=\sum_{\substack{\text{roots of}\\ B(v)}}\text{residue of}\left[\frac{A(v)}{B(v)}\frac{1-\epsilon^{-h(s-v)}}{(s-v)[1-\epsilon^{-T(s-v)}]}\right]\text{ at the roots of } B(v) \tag{13}$$

A special case arises when $R(s)$ has simple poles only,[6,13] then equation 13 reduces to

$$R_p(s)=\sum_{n=1}^{N}\frac{A(v_n)}{B'(v_n)}\frac{1-\epsilon^{-h(s-v_n)}}{(s-v_n)[1-\epsilon^{-T(s-v_n)}]} \tag{14}$$

where

$$B'(v_n)=\left.\frac{dB(v)}{dv}\right|_{v=v_n} \tag{15}$$

and $v_1, v_2 \ldots v_N$ are the simple roots of $B(v)$. Equations 13 and 14 indicate that the p-transform of any function is composed of functions ϵ^{-hs}, ϵ^{-Ts}, and s. Once the transform of the sampler output is found, the usual continuous system procedure can be used to find the final response. To illustrate, consider the system shown in Fig. 4. Let the input be

$$r(t)=1 \text{ (unit step)} \tag{16}$$

$$\text{so that } R(s)=\frac{1}{s} \tag{17}$$

then

$$R_p(s)=\mathrm{P}\left[\frac{1}{s}\right]=\frac{1}{2\pi j}\int_\Gamma\frac{1}{v}\,\frac{1-\epsilon^{-h(s-v)}}{(s-v)[1-\epsilon^{-T(s-v)}]}\,dv \tag{18}$$

$R(v)$ has one simple pole at $v=0$. Hence relation 14 is used to evaluate $R_p(s)$ giving

$$R_p(s)=\frac{1-\epsilon^{-hs}}{s(1-\epsilon^{-Ts})} \tag{19}$$

therefore $C(s)$, the Laplace transform of the output, is

$$C(s)=R_p(s)\;G(s)=\frac{1-\epsilon^{-hs}}{1-\epsilon^{-Ts}}\;\frac{1}{s(s+a)} \tag{20}$$

Table I gives the p-transform of various functions of s. The table, with the aid of partial fraction expansion can be used to determine the p-transform of functions commonly encountered.

Determination of Inverse p-Transform

From equations 13 and 14, and Table I, it is evident that a term in the form of

$$\frac{1-\epsilon^{-h(s+b)}}{1-\epsilon^{-T(s+b)}} \tag{21}$$

is common to all p-transforms for the periodic case being considered. This function has an infinite number of simple poles at

$$s=-b\pm j\frac{2k\pi}{T} \qquad k=1,2,3\ldots \tag{22}$$

Because of these poles, attempts to use the inverse Laplace integral to obtain the time function leads to an infinite series. To avoid this, a relation between the inverse Laplace transform and the inverse modified z-transform has been developed and utilized to give the inverse p-transform in a closed form. Most common p-transforms and their inverses are shown in Table II. The relation between the inverse Laplace transform and the inverse modified z-transform, together with the technique for deriving the inverse p-transform are explained in the Appendix.

Although no knowledge of the modified z-transform is necessary for the use of Table II, a thorough understanding of the notations used is essential. The inverse p-transform is expressed as a function of t,n,m,T, and $d(-h)$, where $d(-h)$

Vol. 75
Pt. 1
1956

Table I. p-Transform Pairs

No.	$F(s)$	$F_p(s)$
1	a_0	a_0
2	$\frac{a_0}{s}$	$\frac{a_0[1-\epsilon^{-hs}]}{s[1-\epsilon^{-Ts}]}$
3	$\frac{a_0}{(s+\gamma)}$	$\frac{a_0[1-\epsilon^{-h(s+\gamma)}]}{(s+\gamma)[1-\epsilon^{-T(s+\gamma)}]}$
4	$\frac{a_0}{s(s+\gamma)}$	$\frac{a_0}{\gamma}\left[\frac{1-\epsilon^{-hs}}{s(1-\epsilon^{-Ts})}-\frac{[1-\epsilon^{-h(s+\gamma)}]}{(s+\gamma)[1-\epsilon^{-T(s+\gamma)}]}\right]$
5	$\frac{a_1s+a_0}{s(s+\gamma)}$	$\frac{1}{\gamma}\left[\frac{a_0(1-\epsilon^{-hs})}{s(1-\epsilon^{-Ts})}-\frac{(a_0-a_1\gamma)[1-\epsilon^{-h(s+\gamma)}]}{(s+\gamma)[1-\epsilon^{-T(s+\gamma)}]}\right]$
6	$\frac{a_2s^2+a_1s+a_0}{s(s+\alpha)(s+\gamma)}$	$\left\{\frac{a_0}{\alpha\gamma}\frac{(1-\epsilon^{-hs})}{s(1-\epsilon^{-Ts})}+\frac{(a_2\alpha^2-a_1\alpha+a_0)}{\alpha(\alpha-\gamma)}\frac{[1-\epsilon^{-h(s+\alpha)}]}{(s+\alpha)[1-\epsilon^{-T(s+\alpha)}]}-\frac{(a_2\gamma^2-a_1\gamma+a_0)}{\gamma(\alpha-\gamma)}\frac{[1-\epsilon^{-h(s+\gamma)}]}{(s+\gamma)[1-\epsilon^{-T(s+\gamma)}]}\right.$
7	$\frac{1}{s^2}$	$\left[\frac{T\epsilon^{-Ts}(1-\epsilon^{-hs})}{s(1-\epsilon^{-Ts})^2}+\frac{1-\epsilon^{-hs}}{s^2(1-\epsilon^{-Ts})}-\frac{h\epsilon^{-hs}}{s(1-\epsilon^{-Ts})}\right]$
8	$\frac{1}{s^N}$	$\frac{(-1)^{N-1}}{(N-1)!}\frac{d^{N-1}}{ds^{N-1}}\left[\frac{1-\epsilon^{-hs}}{s(1-\epsilon^{-Ts})}\right]$
9	$\frac{1}{(s+\gamma)^2}$	$\left\{\frac{T\epsilon^{-T(s+\gamma)}[1-\epsilon^{-h(s+\gamma)}]}{(s+\gamma)[1-\epsilon^{-T(s+\gamma)}]^2}+\frac{1-\epsilon^{-h(s+\gamma)}}{(s+\gamma)^2[1-\epsilon^{-T(s+\gamma)}]}-\frac{h\epsilon^{-h(s+\gamma)}}{(s+\gamma)[1-\epsilon^{-T(s+\gamma)}]}\right\}$
10	$\frac{1}{(s+\gamma)^N}$	$\frac{(-1)^{N-1}}{(N-1)!}\frac{d^{N-1}}{du^{N-1}}\left[\frac{1-\epsilon^{-hu}}{u(1-\epsilon^{-Tu})}\right]_{u=(s+\gamma)}$
11	$\frac{a_1s+a_0}{s^2(s+\gamma)}$	$\left\{\frac{(a_0-a_1\gamma)}{\gamma^2}\frac{[1-\epsilon^{-h(s+\gamma)}]}{(s+\gamma)[1-\epsilon^{-T(s+\gamma)}]}+\frac{(a_1\gamma-a_0)}{\gamma^2}\frac{(1-\epsilon^{-hs})}{s(1-\epsilon^{-Ts})}+\frac{a_0}{\gamma}\left[\frac{T\epsilon^{-Ts}(1-\epsilon^{-hs})}{s(1-\epsilon^{-Ts})^2}+\frac{1-\epsilon^{-hs}}{s^2(1-\epsilon^{-Ts})}-\frac{h\epsilon^{-hs}}{s(1-\epsilon^{-Ts})}\right]\right\}$
12	$\frac{a_1s+a_0}{[(s+\alpha)^2+\beta^2]}$	$\frac{1}{2j\beta}\left\{\frac{[(a_0-a_1\alpha)+ja_1\beta][1-\epsilon^{-h(s+\alpha-j\beta)}]}{(s+\alpha-j\beta)[1-\epsilon^{-T(s+\alpha-j\beta)}]}-\frac{[(a_0-a_1\alpha)-ja_1\beta][1-\epsilon^{-h(s+\alpha+j\beta)}]}{(s+\alpha+j\beta)[1-\epsilon^{-T(s+\alpha+j\beta)}]}\right\}$

$Kf(t)$ denotes a delay of h seconds along the positive time axis operating on the function $Kf(t)$. It is identical to the Laplace notation, $u(t-h)\,Kf(t-h)$; $d(-h)$ is used primarily as a shorthand notation, its meaning illustrated in Fig. 5. Thus

$$d(-h)f(t)=u(t-h)f(t-h) \tag{23}$$

The time t is related to n,m, and T by equation

$$t=(n-1+m)T \tag{24}$$

where

$t\geq 0$
T = period of sampler in seconds
$n=1, 2, 3 \ldots$
= number of sampling periods from zero time. (At $t=0$, n can be zero on condition $m=1$; see equation 24.)
m = a numerical variable of such a form that at all times $0\leq m\leq 1$

Since n is always an integer, then m must be chosen to satisfy relation 24 corresponding to the instant of time t. At the sampling instants in general, it does not matter whether the value $m=0$, or $m=1$ is chosen to satisfy equation 24, but if there are discontinuities in the inverse time function at these instances, then substitution of $m=0$ gives the upper limit and $m=1$ the lower limit. Mathematically this may be stated as

$$\lim_{m\to 0} c(t)\Big|_{t=(n-1+m)T}=c(n-1)T^{+} \tag{25A}$$

$$\lim_{m\to 1} c(t)\Big|_{t=(n-1+mT)}=c(n-1)T^{-} \tag{25B}$$

As t increases n also increases, in the steady state, all terms having ϵ^{-nT} and ϵ^{-t} as multipliers approach zero. Thus these terms form the transient components of the time functions. Inspection of Table II shows that in all cases the inverse time function is composed of a transient term plus a steady-state term. The steady-state condition is readily obtained by letting the terms multiplied by ϵ^{-nT} and ϵ^{-t} assume zero value.

To illustrate the procedure the inverse of equation 20, for the output of the system shown in Fig. 4 is calculated. The output time function $c(t)$ is given by

$$c(t)=\mathrm{P}^{-1}[R_p(s)G(s)]$$
$$=\mathrm{P}^{-1}\left[\frac{1-\epsilon^{-hs}}{1-\epsilon^{-Ts}}\,\frac{1}{s(s+a)}\right] \tag{26}$$

The form is similar to relation 5 of Table II with a_1 equated to zero. Let $a=1$, $T=0.5$ second. If these are substituted in relation 5 of Table II the inverse time function can be written as

$$c(t)=[1-d(-h)]\left\{n-\frac{1}{(\epsilon^{0.5}-1)}[\epsilon^{-0.5(m-1)}-\epsilon^{-t}]\right\} \tag{27}$$

Substituting

$$A=\frac{1}{\epsilon^{0.5}-1}\epsilon^{-0.5(m-1)} \tag{28}$$

$$B=\frac{1}{\epsilon^{0.5}-1}\epsilon^{-t} \tag{29}$$

the time function can be written as

$$c(t)=[1-d(-h)]\ (n-A+B) \tag{30}$$

Table III shows the procedure used for evaluating equation 30 for an assumed pulse width of $h=0.2$ second.

From equation 30 and Fig. 5 it is seen that, once the function $(n-A+B)$ is evaluated, a relative displacement and subtraction can yield the output for any value of pulse width h. This procedure is illustrated graphically in Fig. 6 for $h=0.35$ second and analytically in Table III for $h=0.2$ second. The time function $c(t)$ is plotted in Fig. 7 for various values of h including $h=T=0.5$, the continuous case. (See the section "Limiting Case of p-Transform.")

Other Uses of Inverse p-Transform

From the preceding discussion and examples it is clear that the p-transform is a special form of the Laplace transform. Hence it is evident that

$$r_p(t)=\mathrm{P}^{-1}[R_p(s)]_{t=(n-1+m)T}=\mathfrak{L}^{-1}[R_p(s)] \tag{31}$$

Despite this identity, the process of taking the inverses are markedly different.

$\mathrm{P}^{-1}[R_p(s)]$

denotes a completely different operation on $R_p(s)$ as compared to

$\mathfrak{L}^{-1}[R_p(s)]$

The final form of the time functions $r_p(t)$ also differs greatly in the two cases; they yield the same values only after numerical substitutions.

As pointed out previously (see the section "Determination of Inverse p-Transform") the inverse Laplace leads to an infinite series for the time function. Let this be written as

$$\mathfrak{L}^{-1}[R_p(s)]=x(t)+\text{infinite series in }(t) \tag{32}$$

where $x(t)$ is a closed time function (usually the transient component).

The inverse p-transform, on the other hand, results in a closed form of the time function, expressed in terms of t, n, m, and T. Let this be written as

$$\mathrm{P}^{-1}[R_p(s)]_{t=(n-1+m)T}=y(t)\Big|_{t=(n-1+m)T} \tag{33}$$

where $y(t)$ = transient part + steady-state component. But relation 31 shows equa-

Table II. Inverse p-Transform Pairs

No.	$F_p(s)$	Condition	$f_p(t),\ t=(n-1+m)T,\ t\geq 0$
1	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-Ts}]}\frac{a_0}{s}$		$[1-Kd(-h)]a_0n$
2	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-T(s+q)}]}\frac{a_0}{s}$	$q\neq 0$	$[1-Kd(-h)]\frac{a_0(1-\epsilon^{-nqT})}{(1-\epsilon^{-qT})}$
3	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-T(s+\gamma)}]}\frac{a_0}{(s+\gamma)}$		$[1-Kd(-h)]a_0n\epsilon^{-\gamma t}$
4	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-T(s+q)}]}\frac{a_0}{(s+\gamma)}$	$q\neq\gamma$	$[1-Kd(-h)]\frac{a_0}{(\epsilon^{-T(q-\gamma)}-1)}[\epsilon^{\gamma T}\epsilon^{-T(nq+m\gamma)}-\epsilon^{-\gamma t}$
5	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-Ts}]}\frac{(a_1s+a_0)}{s(s+\gamma)}$		$[1-Kd(-h)]\left\{\frac{a_0n}{\gamma}+\frac{(a_1\gamma-a_0)}{\gamma(\epsilon^{T\gamma}-1)}[\epsilon^{-T\gamma(m-1)}-\epsilon^{-\gamma t}]\right\}$
6	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-T(s+\gamma)}]}\frac{(a_1s+a_0)}{s(s+\gamma)}$		$[1-Kd(-h)]\left\{\frac{a_0(1-\epsilon^{-n\gamma t})}{\gamma(1-\epsilon^{-\gamma T})}+\frac{a_1\gamma-a_0}{\gamma}\ n\epsilon^{-\gamma t}\right\}$
7	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-T(s+q)}]}\frac{(a_1s+a_0)}{s(s+\gamma)}$	$q\neq 0$ $q\neq\gamma$	$[1-Kd(-h)]\left\{\frac{a_0(1-\epsilon^{-nqT})}{\gamma(1-\epsilon^{-qT})}+\frac{(a_1\gamma-a_0)}{\gamma(\epsilon^{-T(q-\gamma)}-1)}[\epsilon^{-T[nq+(m-1)\gamma]}-\epsilon^{-\gamma t}]\right\}$
8	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-Ts}]}\frac{(a_1s+a_0)}{s^2}$		$[1-Kd(-h)]\left\{n\left[a_1+\frac{a_0(mT+t)}{2}\right]\right\}$
9	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-T(s+q)}]}\frac{(a_1s+a_0)}{s^2}$	$q\neq 0$	$[1-Kd(-h)[\left\{\frac{(a_1+a_0mT)\epsilon^{-nTq}-[a_1+a_0(T+t)]}{(\epsilon^{-Tq}-1)}+\frac{a_0T(\epsilon^{-nTq}-1)}{(\epsilon^{-Tq}-1)^2}\right\}$
10	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-T(s+\gamma)}]}\frac{(a_1s+a\ )}{(s+\gamma)^2}$		$[1-Kd(-h)]\left\{n\epsilon^{-\gamma t}\left[a_1+\frac{(a_0-a_1\gamma)(mT+t)}{2}\right]\right\}$
11	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-T(s+q)}]}\frac{(a_1s+a_0)}{(s+\gamma)^2}$	$q\neq\gamma$	$[1-Kd(-h)]\left\{\frac{[a_1+mT(a_0-a_1\gamma)]\epsilon^{-T(nq+m\gamma)}-[a_1+(a_0-a_1\gamma)(T+t)]\epsilon^{-\gamma(T+t)}}{(\epsilon^{-Tq}-\epsilon^{-T\gamma})}+\right.$ $\left.\frac{T(a_0-a_1\gamma)[\epsilon^{-T[nq+(m+1)\gamma]}-\epsilon^{-nT\gamma}\epsilon^{-\gamma t}]}{(\epsilon^{-Tq}-\epsilon^{-T\gamma})^2}\right\}$
12	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-T(s+\gamma)}]^2}\frac{1}{(s+\gamma)}$		$[1-Kd(-h)]\frac{n(n+1)}{2}\ \epsilon^{-\gamma t}$
13	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-T(s+q)}]^2}\frac{1}{(s+\gamma)}$	$q\neq\gamma$	$[1-Kd(-h)]\epsilon^{-\gamma mT}\left[\frac{(n+1)\epsilon^{-nTq}}{(\epsilon^{-Tq}-\epsilon^{-T\gamma})}-\frac{\epsilon^{-(n+1)Tq}-\epsilon^{-(n+1)T\gamma}}{(\epsilon^{-Tq}-\epsilon^{-T\gamma})^2}\right]$
14	$\frac{[1-K\epsilon^{-hs}]}{[1-\epsilon^{-T(s+q)}]\cdot}\frac{(a_1s+a_0)}{[(s+\alpha)^2+\beta^2]}$	$\beta\neq 0$	$[1-Kd(-h)]\left\{\frac{[(a_1\alpha-a_0)^2+(a_1\beta)^2]^{1/2}}{\beta[1+\epsilon^{-2T(q-\alpha)}-2\epsilon^{-T(q-\alpha)}\cos\beta T]^{1/2}}\times\right.$ $\left.[\epsilon^{-T[nq+(m-1)\alpha]}\sin(m\beta T+\theta+\phi)+\epsilon^{-\alpha t}\sin(\beta t+\theta+\psi)]\right\}$ where $\theta=\tan^{-1}\frac{a_1\beta}{a_0-a_1\alpha}$ $\phi=\tan^{-1}\left[\frac{\sin\beta T}{\epsilon^{-T(q-\alpha)}-\cos\beta T}\right]$ $\psi=\tan^{-1}\left[\frac{-\sin\beta T}{\epsilon^{T(q-\alpha)}-\cos\ \beta T}\right]$
15	$\frac{[1-(K_1+jK_2)\epsilon^{-hs}]}{[1-\epsilon^{-T(s+\alpha+j\beta)}]}\frac{1}{(s+\gamma)}$	When $\beta=0$ $\alpha\neq\gamma$	$[1-(K_1+jK_2)d(-h)]\left[\frac{\epsilon^{-T[n\alpha+(m-1)\gamma]}\epsilon^{-jnT\beta}-\epsilon^{-\gamma t}}{\epsilon^{-T(\alpha-\gamma)}\epsilon^{-jT\beta}-1}\right]$
16	$\frac{[1-(K_1+jK_2)\epsilon^{-hs}]}{[1-\epsilon^{-T(s+\rho+j\lambda)}]}\frac{1}{(s+\rho+j\lambda)}$		$[1-(K_1+jK_2)d(-h)]n\epsilon^{-\rho t}\epsilon^{-j\lambda t}$
17	$\frac{[1-(K_1+jK_2)\epsilon^{-hs}]}{[1-\epsilon^{-T(s+\alpha+j\beta)}]}\frac{1}{(s+\rho+j\lambda)}$	$\alpha+j\beta$ $\neq$ $\rho+j\lambda$	$[1-(K_1+jK_2)d(-h)]\left\{\left[\frac{\epsilon^{-T[n\alpha+(m-1)\rho]}\epsilon^{-jT[n\beta+(m-1)\lambda]}-\epsilon^{-\rho t}\epsilon^{-j\lambda t}}{\epsilon^{-T(\alpha-\rho)}\epsilon^{-jT(\beta-\lambda)}-1}\right]\right\}$

Note: By substituting $n=1$, and $mT=t$ in any of these equations, it can be shown analytically that the inverse time function $f(t)$ simplifies to the inverse Laplace for the continuous system. This is expected since for $0\leq t\leq h$ the pulsed system response is identical to the continuous system response.

tions 32 and 33 to be equivalent; hence

$$y(t)\big|_{t=(n-1+m)T}=x(t)+\text{infinite series in }(t) \tag{34}$$

Therefore

$$\text{infinite series in }(t)=y(t)-x(t) \tag{35}$$

Equation 35 indicates that the inverse p-transform can be used to sum up infinite series; see the Appendix. Usually the infinite series part of equation 32 is the Fourier series expansion of the steady-state solution. This leads to the conclusion that the inverse p-transform can be used to find a closed sum for a periodic time function which can be expanded in a Fourier series. A simple example in the Appendix illustrates the applicability of equation 35.

The inverse p-transform has another useful application. This can be demonstrated by considering the following function

$$\frac{1-\epsilon^{-h(s+a)}}{1-\epsilon^{-T(s+q)}}G(s)$$

$$=\left[\frac{G(s)}{1-\epsilon^{-T(s+q)}}-\frac{\epsilon^{-ah}\epsilon^{-hs}G(s)}{1-\epsilon^{-T(s+q)}}\right] \tag{36}$$

The inverse p-transform of this can in general be written as

$$P^{-1}\left[\frac{1-e^{-h(s+a)}}{1-\epsilon^{-T(s+q)}}G(s)\right]$$

$$=[1-\epsilon^{-ah}d(-h)]f(t) \tag{37}$$

$$=[f(t)-\epsilon^{-ah}u(t-h)f(t-h)]$$

Then, clearly

$$P^{-1}\left[\frac{G(s)}{1-\epsilon^{-T(s+q)}}\right]=f(t)$$

$$=\mathcal{L}^{-1}\left[\frac{G(s)}{1-\epsilon^{-T(s+q)}}\right] \tag{38}$$

Thus, effectively, Table II can be considered to be an inverse Laplace table for the functions involving both ϵ^{Ts} and s. Hence Table II can be used to solve general problems arising in Laplace transformations quite independent of whether the sampler is present or not.

To illustrate, consider the effect of applying the periodic saw-toothed voltage of Fig. 8 to a series *R-L* (resistance-inductance) circuit. Assuming zero initial conditions, the relation between the Laplace transforms of the current and voltage may be written as

$$I(s)=\frac{E(s)}{L\left(s+\frac{R}{L}\right)} \tag{39}$$

where

$$I(s)=\mathcal{L}[i(t)] \tag{40}$$

$$E(s)=\mathcal{L}[e(t)]$$

$$=K\left[\frac{(1-\epsilon^{-hs})}{(1-\epsilon^{-Ts})s^2}-\frac{h\epsilon^{-hs}}{s(1-\epsilon^{-Ts})}\right] \tag{41}$$

Substituting equation 41 in equation 39, and simplifying, gives

$$I(s)=\frac{K}{R}\left[\frac{(1-\epsilon^{-hs})}{(1-\epsilon^{-Ts})}\frac{1}{s^2}-\frac{\left[1+\left(\frac{R}{L}h-1\right)\epsilon^{-hs}\right]}{(1-\epsilon^{-Ts})}\frac{1}{s\left(s+\frac{R}{L}\right)}\right] \tag{42}$$

Let

$$\left.\begin{aligned}&R=1\\&L=2\text{ henrys}\\&\frac{R}{L}=0.5\\&h=0.4\text{ second}\\&T=1\text{ second}\\&K=1\end{aligned}\right\} \tag{43}$$

From these and relations 8 and 5 of Table II, the inverse of equation 42 may be written as

$$i(t)=\left\{[1-d(-0.4)]\,n\frac{(m+t)}{2}-[1-0.8d(-0.4)]\left[\frac{n}{0.5}-\frac{1}{0.5(\epsilon^{0.5}-1)}(\epsilon^{-0.5(m-1)}-\epsilon^{-0.5t})\right]\right\} \tag{44}$$

Let

$$A=n\frac{(m+t)}{2} \tag{45}$$

$$B=\frac{n}{0.5}=2n \tag{46}$$

$$C=\frac{1}{0.5(\epsilon^{0.5}-1)}\epsilon^{-0.5(m-1)} \tag{47}$$

$$D=\frac{1}{0.5(\epsilon^{0.5}-1)}\epsilon^{-0.5t} \tag{48}$$

Then $i(t)$ may be written as

$$i(t)=\{[1-d(-0.4)]A-[1-0.8d(-0.4)](B-C+D)\} \tag{49}$$

Table IV shows the procedure used for evaluating equation 49. The resulting $i(t)$ is plotted in Fig. 9.

The foregoing example illustrates the use of Table II for solving problems involving periodic time functions. The advantages in the use of this table are that the solution is exact, avoiding tedious summation of infinite series terms.

Limiting Cases of p-Transform

The *p*-transform must satisfy certain known limiting conditions. It is interesting to study these limiting cases and compare the results with those physically expected.

There are four possible limits:

Case A: $h=0$ with $u_p(t)$ having any finite amplitude K, Fig. 10(A).

Case B: $h=0$ but $u_p(t)$ becomes unit impulses (i.e., has infinite amplitude and area = unity), Fig. 10(B).

Case C: $h=T$, $u_p(t)$ becomes unit step Fig. 10(C).

Case D: h is small compared to time constants of the system Fig. 10(D).

Case A

From Fig. 10(A): $h=0$, $u_p(t)$ has finite amplitude K. This is a trivial case. Under these conditions, the Laplace transform of the pulse function can be written as[14]

$$\mathcal{L}[u_p(t)]=\frac{\int_0^0 K\epsilon^{-st}\,dt}{1-\epsilon^{-Ts}}=0 \tag{50}$$

Thus the sampler output is zero, and the system remains at rest. This is to be expected since under these conditions the system does not receive any energy to cause a change.

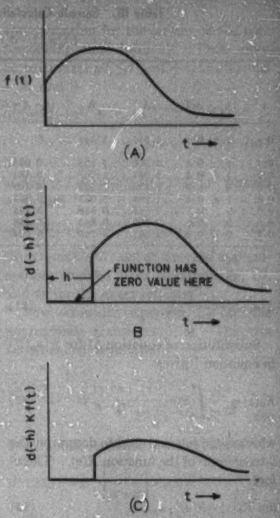


Fig. 5. Illustrating meaning of $d(-h)f(t)$

Case B

From Fig. 10(B): $h=0$, $u_p(t)$ becomes unit impulses (unit area). Then

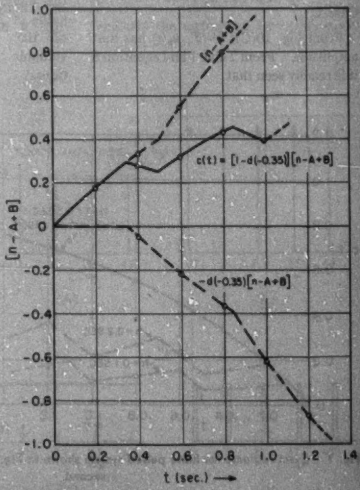


Fig. 6. Graphic method for evaluating output $c(t)$

Table III. Sample Calculation for Evaluating Equation 30

$c(t)=[1-d(-h)](n-A+B)$

t	n	mT	A	B	$(n-A+B)$	$d(-0.2)\times(n-A+B)$	$c(t)$ $h=0.2$	$c(t)$ $h=T=0.5$	$c(t)=\mathcal{L}^{-1}\left[\frac{1}{s(s+1)}\right]$
0	1	0	2.541	1.541	0	↑ $h=0.2$	0	0	0
0.1	1	0.1	2.300	1.395	0.095	↓	0.095	0.095	0.095
0.2	1	0.2	2.081	1.262	0.181	0	0.181	0.181	0.181
0.3	1	0.3	1.883	1.142	0.259	0.095	0.164	0.259	0.259
0.4	1	0.4	1.704	1.033	0.329	0.181	0.148	0.329	0.330
0.5	2	0	2.541	0.935	0.394	0.259	0.135	0.394	0.393
0.6	2	0.1	2.300	0.846	0.546	0.329	0.217	0.451	0.451
0.7	2	0.2	2.081	0.765	0.684	0.394	0.290	0.503	0.503

$$\mathcal{L}[u_p(t)]=\frac{1}{1-\epsilon^{-Ts}} \tag{51}$$

Substitution of equation 51 for $\mathcal{L}[u_p(t)]$ in equation 7 gives

$$R_p(s)=\frac{1}{2\pi j}\int_\Gamma R(v)\frac{1}{1-\epsilon^{-T(s-v)}}\,dv \tag{52}$$

which is the relation used to determine the z-transform[6] of the function $R(v)$. Therefore

$$\lim_{\substack{h\to 0\\ K\to\infty\\ hK\to 1}} R_p(s)=R(z)\big|_{z=\epsilon^{sT}} \tag{53}$$

where K is the amplitude of $u_p(t)$.

This case occurs in problems where mathematically the behavior of the system can be formulated in terms of true impulses; see also case D. In such cases tables of z-transform,[6] modified z-transform[1,7] and their inverse, together with Table II, are suited for finding a solution.

Case C

From Fig. 10(C): $h=T$, $u_p(t)$ has unit amplitude. From Table I and equation 8 it is readily seen that

$$\lim_{h\to T} R_p(s)=R(s) \tag{54}$$

This is clearly an expected result. From equations 54 and 31 it follows that

$$\lim_{h\to T} \mathrm{P}^{-1}[R_p(s)]=\mathcal{L}^{-1}[R(s)] \tag{55}$$

This, although true, is not obvious from the inverse tables. This is because of the particular method used to evaluate the inverse p-transform. However, it is possible to show both analytically (see note in Table II) and numerically, that relation 55 is satisfied by any one of the p-transform pairs listed in Table II. For example, referring to Fig. 4, and from equations 20 and 30, it is to be expected that

$$c(t)=[1-d(-T)]\,(n-A+B)=\mathcal{L}^{-1}\left[\frac{1}{s(s+1)}\right] \tag{56}$$

Table III, together with Fig. 7, shows that this identity is satisfied for the assumed value of $T=0.5$ second. This property of the p-transformation provides a valuable check for numerical calculations.

Case D

From Fig. 10(D): h, small compared to the time constants of the system. This case can be best illustrated by means of an example: Consider the system shown in Fig. 4, with a single time constant $\tau=1/a$. The output is given by equation 20 as

$$C(s)=\frac{1-\epsilon^{-hs}}{(1-\epsilon^{-Ts})s}\,\frac{1}{(s+a)} \tag{20}$$

For $t\geq h$, the numerator can be expanded in powers of hs and assuming, for the time being, that h is sufficiently small so that (h^2s^2) and higher powers can be neglected then $C(s)$ may be closely approximated by

$$C_1(s)=\frac{h}{1-\epsilon^{-Ts}}\,\frac{1}{(s+a)}=\left[\frac{hz}{z-1}\,\frac{1}{(s+a)}\right]_{z=\epsilon^{sT}} \tag{57}$$

where $C_1(s)$ is an approximate transform. Equation 57 is the z-transform approximation used for solving such a system. Comparison of equations 20 and 57 shows that, effectively, the finite pulses in equation 20, with an area hK ($K=1$, unit step input) have been replaced by impulses of equal area, and then multiplied by system transfer function to obtain the approximate transform $C_1(s)$. It is the purpose of the following section to derive the inverse of equation 20 in order to compare it with the approximate inverse obtained from equation 57, and hence show the extent of the approximation caused by the use of equation 57.

For clarity it is advisable to divide the time axis into four regions, corresponding to the four operations of the sampler during each cycle. Let these be

1. $t=nT$, $n=0, 1, 2, 3 \ldots$, the actual sampling instants.
2. $nT<t<nT+h$, the time during which the sampler is definitely closed.
3. $t=nT+h$, the actual instants at which the sampler is transferred from closed to open position.
4. $nT+h<t<(n+1)T$, the time during which the sampler is definitely open.

Regions 1 and 3 are the boundary points.

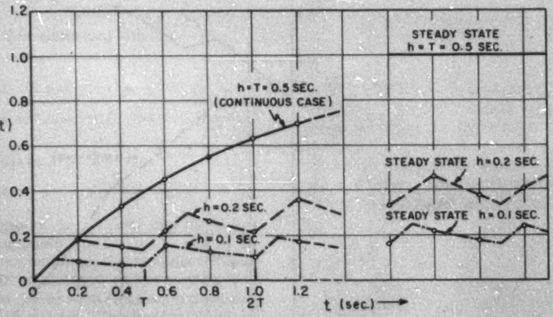


Fig. 7. Exact response for finite pulsed system shown in Fig. 4: $R(s)=1/s$, $a=1$, and $T=0.5$ second

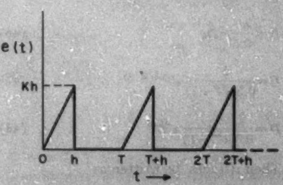


Fig. 8. Periodic saw-toothed voltage

Table IV. Sample Calculation for Evaluating Equation 49

			a	b	e				g	k	l	
t	n	mT=m	A	d(−0.4) A	(a−b)	B	C	D	(B−C+D)	0.8d(−0.4) (B−C+D)	g−k	i(t) e−1
0	1	0	0	0	0	2	5.083	3.083	0	0	0	0
0.4	1	0.4	0.4	0	0.4	2	4.162	2.524	0.362	0	0.362	0.038
0.8	1	0.8	0.8	0.4	0.4	2	3.407	2.067	0.660	0.290	0.370	0.030
1.2	2	0.2	1.4	0.8	0.6	4	4.599	1.692	1.093	0.528	0.565	0.035
1.6	2	0.6	2.2	1.4	0.8	4	3.766	1.385	1.619	0.874	0.745	0.055
2	3	0	3.0	2.2	0.8	6	5.083	1.134	2.051	1.295	0.756	0.044
2.4	3	0.4	4.2	3.0	1.2	6	4.162	0.929	2.767	1.641	1.126	0.074

Equation 20 for the exact transform of the output may be written as

$$C(s)=\frac{1-\epsilon^{-hs}}{s(s+a)}\,[1+\epsilon^{-Ts}+\epsilon^{-2Ts}+\ldots] \quad (58)$$

Let

$$\mathcal{L}^{-1}[R(s)G(s)]=\mathcal{L}^{-1}\left[\frac{1}{s(s+a)}\right]$$
$$=f(t)=\frac{1}{a}\,(1-\epsilon^{-at}) \quad (59)$$

From these the inverse time function can be evaluated term by term as

$$c(t)=\frac{1}{a}\,(1-\epsilon^{-at}) \text{ for } 0\leq t\leq h \quad (60)$$

$$c(t)=\frac{1}{a}\,(\epsilon^{ah}-1)\epsilon^{-at} \text{ for } h\leq t\leq T \quad (61)$$

These may be generalized as follows for $nT\leq t\leq nT+h$ i.e., regions 1, 2, and 3.

$$c(t)=\left\{\left(\frac{\epsilon^{ah}-1}{a}\right)[1+\epsilon^{aT}+\epsilon^{2aT}+\ldots+\epsilon^{(n-1)aT}]\epsilon^{-at}+\frac{1}{a}[1-\epsilon^{-a(t-nT)}]\right\} \quad (62)$$

and for $nT+h\leq t\leq(n+1)T$, i.e., regions 3, 4, and 1.

$$c(t)=\frac{\epsilon^{ah}-1}{a}\,[1+\epsilon^{aT}+\epsilon^{2aT}+\ldots+\epsilon^{anT}]\epsilon^{-at} \quad (63)$$

The output at the sampling instants $t=nT$ can be obtained from equation 62 and is

$$c(nT)=\frac{\epsilon^{ah}-1}{a}\,[\epsilon^{-aT}+\epsilon^{-2aT}:\ldots+\epsilon^{-anT}]+f(0) \quad (64)$$

where

$$f(0)=f(t)\big|_{t=0}$$

(see equation 59.) Equation 64 gives the true output at the sampling instants $t=nT$, since it is derived from the exact transform. When the ratio $ah=h/\tau$, where τ=time constant, is small so that (a^2h^2) and higher powers can be neglected, then $\epsilon^{ah}-1/a$ can be closely approximated by

$$\frac{\epsilon^{ah}-1}{a}\simeq h \quad (65)$$

Substituting equation 65 in 64, a true approximation for the output at times $t=nT$ is obtained. This is

$$c_a(nT)=h[\epsilon^{-aT}+\epsilon^{-2aT}+\ldots+\epsilon^{-anT}]+f(0) \quad (66)$$

Similarly, substitution of relation 65 in equations 62 and 63 results in a true approximation for the output at the corresponding time intervals. These may be written for $nT\leq t\leq nT+h$, i.e., regions 1, 2, and 3

$$c_a(t)=h[1+\epsilon^{aT}+\ldots+\epsilon^{(n-1)aT}]\epsilon^{-at}+\frac{1}{a}[a-\epsilon^{-a(t-nT)}] \quad (67)$$

and for $(nT+h)\leq t\leq(n+1)T$, i.e., regions 3, 4, and 1

$$c_a(t)=h[1+\epsilon^{aT}+\epsilon^{2aT}+\ldots+\epsilon^{anT}]\epsilon^{-at} \quad (68)$$

Equations 66, 67, and 68 are valid approximations derived from equations 62, 63, and 64 for the exact response.

Now, consider equation 57 for the approximate transform $C_1(s)$. This can be expanded as follows

$$C_1(s)=\frac{1}{(s+a)}\,[1+\epsilon^{-Ts}+\epsilon^{-2Ts}+\ldots+\epsilon^{-nTs}+\ldots] \quad (69)$$

The impulse response of the system is

$$\mathcal{L}^{-1}[G(s)]=\mathcal{L}^{-1}\left[\frac{1}{s+a}\right]=g(t)=\epsilon^{-at} \quad (70)$$

Thus the inverse transform of equation 69 representing the approximate inverse time function is

$$c_1(t)=h[1+u(t-T)+u(t-2T)+\ldots]\epsilon^{-at} \quad (71)$$

From this the output at the sampling instants $t=nT$ may be written as

$$c_1(nT)=h[g(0)+\epsilon^{-aT}+\epsilon^{-2aT}+\ldots+\epsilon^{-anT}] \quad (72)$$

where

$g(0)=g(t)\big|_{t=0}=1$ (in this case)

Similarly the output at any other time

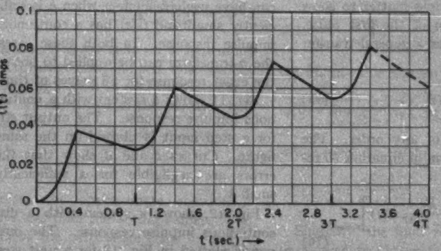


Fig. 9 (above). Plot of current i(t) in the series R/L circuit for the applied saw-toothed voltage

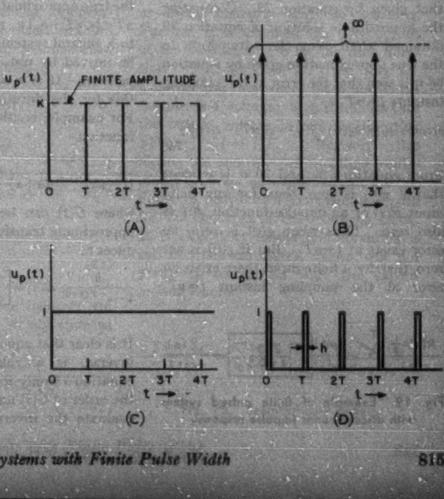


Fig. 10 (right). Limiting cases of pulse function $u_p(t)$

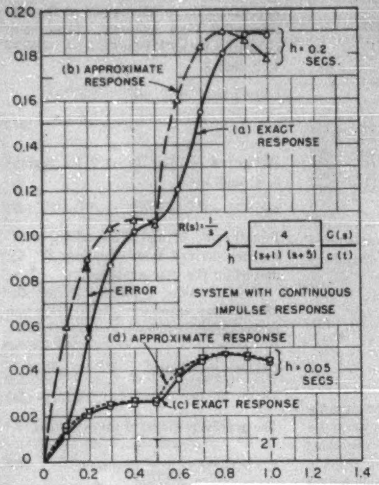


Fig. 11. Response of system with continuous impulse response. Plot shows that errors are small for sufficiently small h

$nT<t<(n+1)T$, i.e., regions 2, 3, and 4 can be written as

$$c_1(t)=h[1+\epsilon^{aT}+\epsilon^{2aT}+\ldots+\epsilon^{anT}]\epsilon^{-at} \tag{73}$$

Equations 72 and 73 are the resulting time functions when equation 20 is approximated by equation 57.

It is known that z-transform can only give information at sampling instants $t=nT$. Hence, if the z-transform is applied to equation 57 to evaluate the response of the system at these instants, the result obtained will be identical to that given by equation 72. Comparing the approximate solution of equation 72 obtained when using the z-transform to the true approximation given by equation 66 it is seen that the error at the sampling instants $t=nT$ is

$$\text{error}\Big|_{t=nt}=hg(0)-f(0)=c_1(nT)-c_2(nT) \tag{74}$$

From equations 59 and 70 it is evident that if $g(0)$ is zero, then for any finite input $R(s)$ or an impulse function, $f(0)$ is also zero. Thus when $g(0)$ is zero, no error exists at $t=nT$. But if $g(0)$ is not zero, then for a finite input there exists an error at the sampling instant $t=nT$.

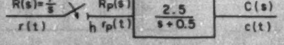


Fig. 12. Example of finite pulsed system with discontinuous impulse response

Hence it can be stated that the use of z-transform (equation 57) for determining the solution of finitely pulsed system at sampling instants, $t=nT$ is not correct, leading to wrong results, unless the system transfer function $G(s)$ has a continuous impulse response, i.e., the degree of its denominator in s is two or more higher than the degree of its numerator.

Comparison of equation 73 with equations 67 and 68 also shows that the time domain solution derived from the approximate transform does not correspond with the true approximation during the interval $nT<t<nT+h$, i.e., region 2. This applies to a general system of any order and can be proved by use of partial fraction expansion. It is a simple matter to show this for the duration of the first pulse. For example, consider a general transfer function

$$\frac{1-\epsilon^{-h(s+a)}}{[1-\epsilon^{-T(s+a)}](s+a)}G(s) \tag{75}$$

where $G(s)$ can be of any order. The approximate transform of equation 75 reduces to

$$\frac{h}{1-\epsilon^{-T(s+a)}}G(s)=\left[\frac{hz}{z-\epsilon^{-aT}}G(s)\right]_{z=\epsilon^{sT}} \tag{76}$$

It is clear that equation 76 (and hence its inverse) is a valid approximation of equation 75 only for $t\geq h$, irrespective of the order of $G(s)$ and the method used to evaluate the inverse. From these considerations a statement can be formulated as follows:

Consider a system with transfer function $G(s)$, the input to which is a sequence of finite pulses of width h and period T. When h is small compared to the time constants of the systems, the output can be approximated by replacing the finite pulses by impulses of equivalent area. In general two cases arise.

Case A: $G(s)$ has discontinuous impulse response. In this case, the solution is a true approximation for the time duration $nT+h\leq t<(n+1)T$, i.e., regions 3 and 4. This does not include the sampling instants $t=nT$. Hence the z-transform approximation completely fails in this case.

However, the modified z-transform can be used to obtain the output over the duration $nT+h\leq t<(n+1)T$. And by a limiting process at $(n+1)T$. (The modified z-transform gives results between sampling instants as well as the sampling instants. At a discontinuity it is possible to obtain both the upper and lower limits.)

Case B: $G(s)$ has continuous impulse response. In this case the approximate solution is valid for the time duration $nT+h\leq t\leq(n+1)T$, i.e., regions 3, 4, and 1. This includes the sampling instants $t=nT$. Here, both the z-transform and modified z-transform can be used to obtain a true approximate solution.

From these it is seen that the solutions derived from the approximate transforms, equations 57 and 76, are not a true approximation of the system output during the interval $nT<t<nT+h$, i.e., region 2, irrespective of the order of $G(s)$ and the method used to evaluate the inverse. This means that when the z-transform approximation is used, the true behavior of the system during the interval $nT<t<nT+h$ is not known. This, however, can have negligible effect when h is sufficiently small and $G(s)$ has a continuous impulse response. In both these cases, the p-transform can be applied to obtain the exact response for all values of time.

Fig. 11 shows a system with a continuous impulse response. The output $c(t)$ for a step input is plotted on the same figure. Curves c and d show that the errors are negligible for a sufficiently small h.

Fig. 12, shows a system with a discontinuous impulse response. The output is plotted in Fig. 13 for a unit step input, and assumed values of $h=0.2$ second and T=0.5 second. Curve a is the exact time function calculated from the inverse p-transform. Curve b is the approximate response calculated from the

Vol. 7
Pt. 1
1956

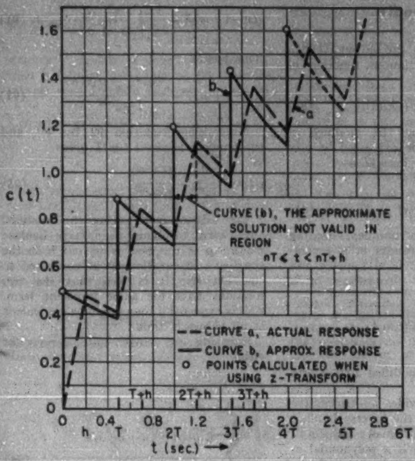


Fig. 13. Response of system shown in Fig. 12 to step input

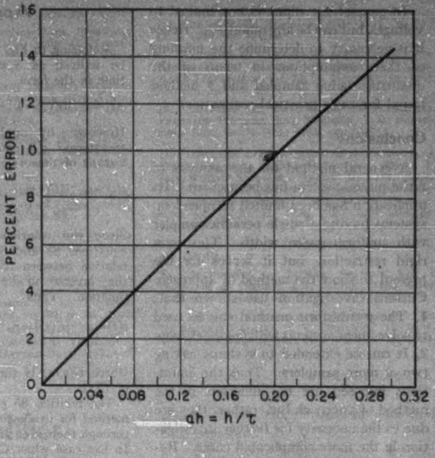


Fig. 14. Percentage error as function of ah for single time constant system

approximate transform. Comparison of the two curves shows the extent of the error which can be caused both at the sampling instants, and for the duration $nT<t<nT+h$ where the approximate transform is not valid.

In the foregoing the degree of approximation depends on the ratio of h to the time constants of the system. From equations 63, 65, and 68 it is seen that for the single time constant system of Fig. 4, with unit step input the approximations are due to substitution of

$$\frac{\epsilon^{ah}-1}{a}=h \qquad (65)$$

From this the percentage error can be expressed as

$$\text{per cent of error}=\left[1-\frac{ah}{\epsilon^{ah}-1}\right]\times 100 \qquad (77)$$

Equation 77 shows that for a step input the percentage error is fixed, independent of sampling period T, and time t. For a stable system (poles in the negative half of s-plane)

$$\frac{\epsilon^{ah}-1}{a}>h \qquad (78)$$

therefore the approximate solution is always less than the actual values. Fig. 14 shows the percentage error for a system with single time constant $\tau=1/a$ as a function of ah. It is seen that for an error of less than five per cent, ah must be less than or equal to one-tenth. For systems with more than one time constant the total error varies with time, as shown in Fig. 11.

Fig. 12 shows a single time constant system with $\tau=2$ seconds. Its output is plotted in Fig. 13 assuming a pulse width $h=0.2$, $T=0.5$, and a unit step input. Inspection shows that in its region of validity $(nT+h\leq t<(n+1)T)$ the approximate solution is less than the actual output, and the percentage error is nearly 5 per cent corresponding to $ah={}^1/_{10}$.

Initial and Final Values

Since p-transform is only a particular form of the Laplace transform, all operations valid for the Laplace transformation are equally applicable to p-transforms. For example, given a function

$$C_p(s)=\frac{1-\epsilon^{-h(s+a)}}{[1-\epsilon^{-T(s+a)}](s+a)}G(s) \qquad (79)$$

the initial value is given by

$$c_p(0)=\lim_{s\to\infty}\frac{s[1-\epsilon^{-h(s+a)}]}{[1-\epsilon^{-T(s+a)}](s+a)}G(s) \qquad (80)$$

which is equivalent to

$$c_p(0)=\lim_{s\to\infty}\frac{sG(s)}{(s+a)} \qquad (81)$$

Similarly, the final value theorem can be applied to obtain the steady state as the following

$$\lim_{t\to\infty}c_p(t)=\lim_{s\to 0}\frac{s[1-\epsilon^{-h(s+a)}]}{[1-\epsilon^{-T(s+a)}](s+a)}G(s) \qquad (82)$$

However, in most open loop cases, the final steady state is oscillating (see Fig. 7), hence the final value theorem is not applicable.

Practical Realization of Sampler

To correlate the theoretical calculations with practical results it is necessary that elements of the system be closely approximated by their idealized mathematical descriptions.

In the preceding discussions an ideal sampler has been assumed. That is, it is assumed that the sampler can make or break instantly with no sparking. Further, it is assumed that the output of the sampler is zero for the time duration $nT+h\leq t\leq(n+1)T$; see Fig. 2. Consider the case where the input to the sampler is a voltage; then it follows that the output voltage is zero for duration $nT+h\leq t\leq(n+1)T$. Effectively, this means the sampler (or sampling equipment) must act as a short circuit for this duration, providing a short-circuited path for the elements of the following block. It is clear that unless this condition is satisfied the calculated values will not prove to correspond with the experimental results.

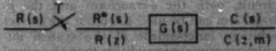


Fig. 15. Sampled-data system

In general the input is not limited to voltages, but can be any quantity. Hence it is necessary to determine the meaning of these assumptions in terms of the quantities being sampled and a method found for their practical realization.

Conclusions

A general method for the analysis of finite pulsed system has been given. Its application has been limited to open-loop systems having a single periodic sampler with uniform pulse width. This is a rigid restriction, but it serves for the present to show the method of approach. Current investigation has shown that: 1. The p-transform method can be used to solve linear systems with feedback loop. 2. It can be extended to systems having two or more samplers. Thus the limitations imposed are not inherent in the method of approach but, rather, they are due to the necessity for further investigation in the more complicated cases. Results of these studies will be published in the near future.

In this paper it has been shown that:

1. The inverse p-transform table can be used to obtain the exact inverse Laplace of functions involving both ϵ^{Ts} and s.
2. Since the Laplace transform of periodic time functions are in the form of $F(s)/(1\mp\epsilon^{-Ts})$, then Table II can be used to obtain the inverse time function. However, any periodic time function has a Fourier series expansion in the time domain. Hence the inverse p-transform table can be used to find a closed form for the infinite series Fourier expansion.
3. It is shown that for a finite input and in the absence of a hold circuit, the z-transform leads to erroneous results unless the system transfer function has a continuous impulse response.
4. Application of p-transform results in the correct response irrespective of the form of the system impulse response. This includes systems whose transfer function have the same order polynomial in the numerator and denominator.
5. When the system transfer function has a continuous impulse response and the pulse width h is small, i.e., it is approximately equal to one tenth to one fifth of the smallest time constant of the system, then the z-transform technique can be used in order to obtain a reasonably approximate solution.
6. When the z-transform approximation is used, the true behavior of the system is not known during the interval in which the sampler is definitely closed.
7. When the input to the system consists of true impulses, or the system has a hold circuit, both the z-transform and the p-transform can be used to obtain the correct solution, independent of the system impulse response, on condition that the actual transfer functions allowing for the existing pulse width are considered.

Appendix

Referring to Fig. 15, the output $c(t)$ can be written as a continuous function of time in the form

$$c(t)=\mathcal{L}^{-1}[C(s)=\mathcal{L}^{-1}[R^*(s)G(s)] \qquad (83)$$

However, by applying the modified z-transform,[1] $c(t)$ can be expressed at any instant of time $t=(n-1+m)T$ as

$$\begin{aligned}c(t)&=z_m{}^{-1}[C(z,m)]\\&=z_m{}^{-1}[R(z)G(z,m)]_{t=(n-1+m)T} \qquad (84)\end{aligned}$$

Since the solution is unique, equating equations 83 and 84 results in a general relation between the inverse Laplace and the inverse modified z-transform of a function. Thus

$$c(t)=\mathcal{L}^{-1}[C(s)]=z_m{}^{-1}[C(z,m)]_{t=(n-1+m)T} \qquad (85)$$

where $C(z,m)$ is the modified z-transform of $C(s)$.

The relation 85 provides an alternative method for evaluating the inverse Laplace through the use of the modified z-transform. In the case when $C(s)$ is a polynomial in s, it is simpler to use the ordinary methods, and evaluate $c(t)$ directly from $C(s)$. But in cases when $C(s)$ is the Laplace transform of a periodic function, i.e., when it is of the form

$$\frac{1}{1\mp\epsilon^{-Ts}}F(s)$$

or has factorable terms in the form of

$$\epsilon^{+Ts}$$

then it becomes advantageous to obtain the inverse time function through the use of the modified z-transform. To illustrate consider the example

$$C(s)=\frac{1}{1-\epsilon^{-Ts}}\,\frac{1}{(s+a)} \qquad (86)$$

the inverse Laplace of this can be written as

$$c(t)=\frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty}\frac{\epsilon^{st}}{(1-\epsilon^{-Ts})(s+a)}\,ds \qquad (87)$$

the integrand has simple poles at $s=-a$ and $s=\pm j2k\pi/T$, $k=0, 1, 2, \ldots$ evaluating the integral at these simple poles gives

$$c(t)=\frac{\epsilon^{-at}}{1-\epsilon^{aT}}+\frac{1}{aT}+\frac{2}{T}\times\sum_{k=1}^{\infty}\frac{a\cos\frac{2\pi k}{T}t+\frac{2\pi k}{T}\sin\frac{2\pi k}{T}t}{a^2+\left(\frac{2\pi k}{T}\right)^2} \qquad (88)$$

Utilizing relation 85, the inverse time function may also be written as

$$c(t)=\mathcal{L}^{-1}[C(s)]=z_m{}^{-1}[C(z,m)]_{t=(n-1+m)T} \qquad (89)$$

where $C(z,m)$ is the modified z-transform of $C(s)$, and is

$$C(z,m)=\frac{z}{z-1}\,\frac{\epsilon^{-amT}}{z-\epsilon^{-aT}} \qquad (90)$$

The inverse of this is given by[1]

$$c(n-1+m)T=\frac{1}{2\pi j}\int_{\Gamma}\frac{z^n}{z-1}\,\frac{\epsilon^{-amT}}{z-\epsilon^{-aT}}\,dz \qquad (91)$$

This yields

$$c(n-1+m)T=\frac{\epsilon^{-amT}-\epsilon^{-aT(n+m)}}{1-\epsilon^{-aT}} \qquad (92)$$

Multiplying equation 92 by ϵ^{aT} and substituting $t=(n-1+m)T$ gives

$$c(t)=\frac{\epsilon^{-at}}{1-\epsilon^{aT}}-\frac{\epsilon^{aT(1-m)}}{1-\epsilon^{aT}} \qquad (93)$$

The advantage of using the modified z-transform to obtain the inverse Laplace transform in this case is evident from the simplicity of equation 93 as compared to equation 88. It is seen that the two solutions have the same transient term. This is expected if they are to be equivalent at all values of time.

Equating the steady-state component yields

$$\frac{\epsilon^{aT(1-m)}}{\epsilon^{aT}-1}=\frac{1}{aT}+\frac{2}{T}\times\sum_{k=1}^{\infty}\frac{a\cos\frac{2\pi k}{T}t+\frac{2\pi k}{T}\sin\frac{2\pi k}{T}t}{a^2+\left(\frac{2\pi k}{T}\right)^2} \qquad (94)$$

from which, by substitution of $t=(n-1+m)T$ is obtained the relation

$$\sum_{k=1}^{\infty}\frac{a\cos 2\pi km+\frac{2\pi k}{T}\sin 2\pi km}{a^2+\left(\frac{2\pi k}{T}\right)^2}=\frac{T[\epsilon^{aT(1-m)}]}{2(\epsilon^{aT}-1)}-\frac{1}{2a} \qquad (95)$$

$$0<m<1$$

Equation 95 is a by-product of the relation 85. The right-hand side is the closed form of the infinite series on the left-hand side. This identity has been checked and found to be correct for various values of a, T, and m.

In general, if $C(s)$ is the Laplace transform of a periodic function, the inverse Laplace is always in the form of an infinite series, but the use of relation 85 enables one to find a closed form for this infinite series, thus there is no reason why $C(s)$ should not be synthesized to give an infinite series whose sum is to be found in the time domain, and then the modified z-transform applied to get this sum.

Equation 93 shows that the inverse time function $c(t)$ can be written as

$$c(t)=\mathcal{L}^{-1}\left[\frac{1}{1-\epsilon^{-Ts}}\,\frac{1}{s+a}\right]=\frac{1}{\epsilon^{aT}-1}[\epsilon^{aT(1-m)}-\epsilon^{-at}] \qquad (96)$$

This identity can be used to evaluate the inverse p-transform of certain functions in a closed form. To illustrate, consider relation 4 of Table II. Let $q=0$, then $F_p(s)$ becomes

$$F_p(s)=\frac{1-k\epsilon^{-hs}}{1-\epsilon^{-Ts}}\,\frac{a_0}{(s+\gamma)} \qquad (97)$$

This may be written as

$$F_p(s) = [1 - k\epsilon^{-hs}]\left[\frac{1}{1-\epsilon^{-Ts}}\,\frac{a_0}{(s+\gamma)}\right] \quad (98)$$

since ϵ^{-hs} denotes a delay of h seconds, then it follows from equation 96 that the inverse time function $f_p(t)$ can be written

$$f_p(t) = [1 - kd(-h)]\left\{\frac{a_0}{\epsilon^{\gamma T}-1}[\epsilon^{\gamma T(1-m)} - \epsilon^{-\gamma t}]\right\} \quad (99)$$

which is the same as that listed in relation 4 of Table II with $q=0$. It is evident from the preceding that the inverse p-transform of other functions can be evaluated in a similar way. Thus, relation 85 can be advantageously applied to obtain the inverse p-transform in a closed form.

References

1. Synthesis and Critical Study of Sampled-Data Control Systems, Eliahu I. Jury. *AIEE Transactions*, vol. 75, pt. II, July 1956, pp. 141–51.

2. The Design of Sampled-Data Feedback Systems, Gladwyn V. Lago, John G. Truxal. *AIEE Transactions*, vol. 73, pt. II, Nov. 1954, pp. 247–53.

3. Sampled-Data Control Systems Studied Through Comparison of Sampling with Amplitude Modulation, William K. Linvill. *Ibid.*, vol. 70, pt. II, 1951, pp. 1779–88.

4. Theory of Servomechanisms, H. M. James, N. B. Nichols, R. S. Phillips. McGraw-Hill Book Company, Inc., New York, N. Y., 1947, chap. 5 by W. Hurewitz.

5. The Analysis of Sampled-Data Systems, J. R. Ragazzini, L. A. Zadeh. *AIEE Transactions*, vol. 71, pt. II, Nov. 1952, pp. 225–32.

6. Analysis and Synthesis of Sampled-Data Control Systems, Eliahu I. Jury. *Ibid.*, vol. 73, pt. I, Sept. 1954, pp. 332–46.

7. The Pulse Transfer Function and Its Application to Sampling Servo-systems, R. H. Barker. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 99, pt. IV, 1952, pp. 302–17.

8. Extension of Conventional Techniques to the Design of Sampled-Data Systems, W. K. Linvill, R. W. Sittler. *Convention Records*, Institute of Radio Engineers, New York, N. Y., pt. I, 1953, pp. 99–104.

9. Additions to z-Transformation Theory for Sampled-Data Systems, Gladwyn V. Lago. *AIEE Transactions*, vol. 73, pt. II, 1954 (Jan. 1955 section), pp. 403–08.

10. Analysis of Errors in Sampled-Data Feedback Systems, Jack Sklansky, J. R. Ragazzini. *Ibid.*, pt. II, May 1955, pp. 65–71.

11. Discrete Compensation of Sampled-Data and Continuous Control Systems, Eliahu I. Jury, William Shroeder. *Report No. 154, Series 60*, Electronics Research Laboratory, University of California, Berkeley, Calif. *AIEE Transactions*, vol. 75, pt. II, 1956 (Jan. 1957 section), pp. 317–25.

12. *Discussion* by G. Farmanfarma of reference 1, pp. 149–50.

13. Transients in Linear Systems, Vol. I (book), M. F. Gardner, J. L. Barnes. John Wiley & Sons, Inc., New York, N. Y., 1953, pp. 275–78.

14. Modern Operational Mathematics in Engineering (book), R. V. Churchill. McGraw-Hill Book Company, Inc., New York, N. Y., 1944, p. 54.

The Series Magnetic Amplifier

Part I—The Four Modes of Operation; Part II—The Mechanisms of Gain and Time Constant

R. C. BARKER
ASSOCIATE MEMBER AIEE

THE THEORETICAL writings of Storm[1-3] on magnetic amplifiers have deservedly received much attention. The references given to his work cite particularly his theory of the series-connected saturable reactor, to be called here the "series amplifier." It is assumed in the following that the reader is familiar with Storm's analysis of the series amplifier with comparatively low and comparatively high control-circuit resistance.

Two questions commonly arise regarding the solutions for extreme values of control resistance. One is "Why should the method of analysis depend upon the value of the control resistance?" and the other, "How does the control voltage 'control' the circuit?" If a generalized analysis were made, without assigning values to the control voltage and control resistance, one set of solutions would result which would apply for extreme and intermediate values of the control-circuit parameters as well.

Part I of this paper presents such a generalized analysis for the steady-state operation of the amplifier. The control voltage and resistance appear as parameters in the solutions. Two kinds of control voltages are considered: the direct voltage and the rectified sine wave. It is shown that the natural classification of amplifier operation is by modes defined by sequences of saturation states of the cores. The problem is to find the restrictions on circuit parameters and the control voltage for the existence of any given mode.

Perhaps the most important fact to note in Part I is that all the solutions are taken from one equation. The solutions for the four modes require some algebraic manipulations. But the only real theory involved is contained in a well-known average-flux theorem and a symmetry theorem, and the simple equation they imply. Part II shows that this one equation also contains a powerful physical picture of the mechanisms of gain and time constant.

Series Amplifier with D-C Control Voltage

In the following analysis, amplifier operation is divided into four distinct modes. These modes are distinguished by sequences of saturation states which occur within a cycle of the supply voltage. There are four of these saturation states: 1. neither core saturated, 2. core a saturated, 3. core b saturated, and 4. both cores saturated.

In each of the four saturation states, the circuit is represented by two linear voltage equations. Solutions for the circuit voltages, currents, and fluxes are easily obtained in each of the four saturation states. Within a given cycle, a succession of saturation states may exist, each for a finite angular interval. The problem remaining is to join these intervals to form a sequence of saturation states to meet certain requirements of symmetry and flux variation for the steady state.

CIRCUIT EQUATIONS FOR THE FOUR SATURATION STATES

The circuit to be considered is that of Fig. 1. The piecewise linear magnetic characteristics of Fig. 2 are assumed to apply. All windings are assumed to be resistanceless. The source of control signal is a direct voltage with an internal resistance R_c as shown in Fig. 1. The circuit consists of two electrical meshes coupled together by the two saturable reactors, core a and core b. Consequently, there are two voltage equations and two magnetic equations which can be written for the circuit. The special constraints dictated by a particular saturation

---

Papers 56-747, 56-748, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript, Part I, submitted March 28, 1956; made available for printing May 29, 1956; Part II submitted March 28, 1956; made available for printing June 8, 1956.

R. C. Barker is with the Yale University Department of Electrical Engineering, New Haven, Conn.

This paper is a partial abstract of the dissertation presented by the author to the Graduate School of Yale University, May 1, 1955. The original work was done under the sponsorship of the Wright Air Development Center. The project scientist was Dr. Rolf M. Wundt. Special acknowledgment is due Prof. Albert G. Conrad and Prof. Walter J. Cunningham for their advice in the performance of this research.

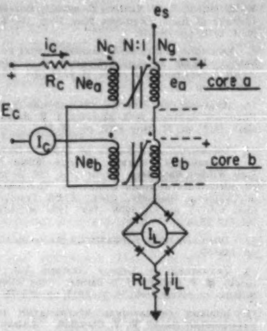


Fig. 1. Series magnetic amplifier

state can be introduced, and the solutions for the equations can be simply obtained for that state.

The voltage equations for the load and control sides of the amplifier are

$$e_s = e_a + e_b + i_L R_L \tag{1}$$

$$E_c = Ne_a - Ne_b + i_c R_c \tag{2}$$

The magnetic equations are written by referring to the assumed magnetic characteristics of Fig. 2

$$F_a = N_c i_c + N_g i_L \tag{3}$$

$$F_b = -N_c i_c + N_g i_L \tag{4}$$

For neither core saturated, the special constraints are

$$F_a = F_b = 0 \tag{5}$$

Equations 1 and 2 then become

$$e_s = e_a + e_b \qquad E_c = Ne_a - Ne_b$$

The desired solutions for the voltages and currents are obtained by applying the constraints of equation 5 to equations 1, 2, 3, and 4.

$$i_L = i_c = 0 \tag{6}$$

$$e_a = \frac{e_s}{2} + \frac{E_c}{2N} \tag{7}$$

$$e_b = \frac{e_s}{2} - \frac{E_c}{2N} \tag{8}$$

For core *a* saturated, the special constraints are

$$e_a = 0 \tag{9}$$

$$F_b = 0 \tag{10}$$

These constraints are first applied to equations 1 through 4. Equation 2 is then divided by *N* and added to equation 1.

$$e_s + \frac{E_c}{N} = i_L\left(R_L + \frac{R_c}{N^2}\right) \tag{11}$$

This equation brings into evidence the equivalent generator in the load circuit and the equivalent resistance through which the load current flows. This equivalent resistance is defined as

$$R_L' = R_L + \frac{R_c}{N^2} \tag{12}$$

The desired solutions for voltages and currents in the second saturation state then become

$$i_L = \left(e_s + \frac{E_c}{N}\right)\frac{1}{R_L'} \tag{13}$$

$$i_c = \left(e_s + \frac{E_c}{N}\right)\frac{1}{NR_L'} \tag{14}$$

$$e_b = e_s\left(1 - \frac{R_L}{R_L'}\right) - \frac{E_c}{N}\,\frac{R_L}{R_L'} \tag{15}$$

For core *b* saturated, the special constraints are

$$e_b = 0 \tag{16}$$

$$F_a = 0 \tag{17}$$

If one follows a procedure similar to that for state 2, the desired solutions for voltages and currents in the third saturation state become

$$i_L = \left(e_s - \frac{E_c}{N}\right)\frac{1}{R_L'} \tag{18}$$

$$i_c = -\left(e_s - \frac{E_c}{N}\right)\frac{1}{NR_L'} \tag{19}$$

$$e_a = e_s\left(1 - \frac{R_L}{R_L'}\right) + \frac{E_c}{N}\,\frac{R_L}{R_L'} \tag{20}$$

For both cores saturated, the special constraints are

$$e_a = 0 \tag{21}$$

$$e_b = 0 \tag{22}$$

The solutions for the currents become

$$i_L = \frac{e_s}{R_L} \tag{23}$$

$$i_c = \frac{E_c}{R_c} \tag{24}$$

Equivalent circuits for the four saturation states are shown in Fig. 3.

The solutions for the reactor voltages and currents in the four saturation states are only partial solutions, i.e., the solutions for one state are valid only for the angular interval corresponding to that state. The difficult part of finding the total solutions of the circuit equations is determining the states which appear in sequence and the order in which they occur. A sequence can be assumed, and the equation defining the steady state can be written. Over the range of parameters for which the solution of this equation exists, the assumed sequence is correct. For a given sequence, or mode, the average-flux and symmetry theorems[4] give the integral equation which can be solved for the firing angle as a function of the control voltage. These theorems are briefly stated in the following.

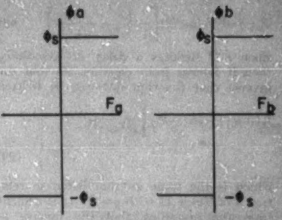


Fig. 2. Assumed magnetic characteristics

Average-flux theorem as it applies to this problem can be stated as follows: In the steady state, the net flux change in each core over any interval equal to the period of the supply frequency is zero. In concise form

$$\int_{\omega t}^{\omega t+2\pi} e_a d\lambda = \int_{t\omega}^{\omega t+2\pi} e_b d\lambda = 0 \tag{25}$$

The symbol λ is a general angular variable of integration.

The symmetry theorem for this circuit in the steady state is

$$e_a(\omega t) = -e_b(\omega t + \pi) \tag{26A}$$

$$e_b(\omega t) = -e_a(\omega t + \pi) \tag{26B}$$

The second integral of equation 25 can be written

$$\int_{\omega t}^{\omega t+\pi} e_b d\lambda + \int_{\omega t+\pi}^{\omega t+2\pi} e_b d\lambda = 0 \tag{27}$$

Equation 26(A) can be substituted into equation 27 with the result for the steady state

$$\int_{\omega t}^{\omega t+\pi} (e_a - e_b) d\lambda = 0 \tag{28}$$

All of the solutions for steady-state firing angle are found by performing the foregoing integration.

THE FOUR MODES OF CIRCUIT OPERATION

The first step in developing the four modes of circuit operation is to assume the sequence of saturation states that was found to exist in the simplified analysis of Storm. In this analysis it was found that twice per cycle state 1 appeared, alternating with intervals of state 2 and state 3. This sequence defines mode I. To illustrate the manner in which the sequence of saturation states defines the mode of operation, the four modes are summarized in Fig. 4.

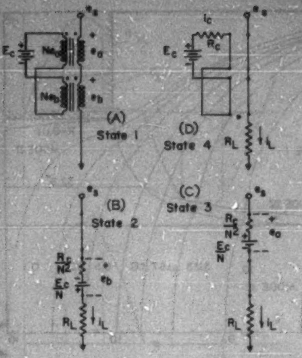


Fig. 3 (left). Equivalent circuits for four saturation states

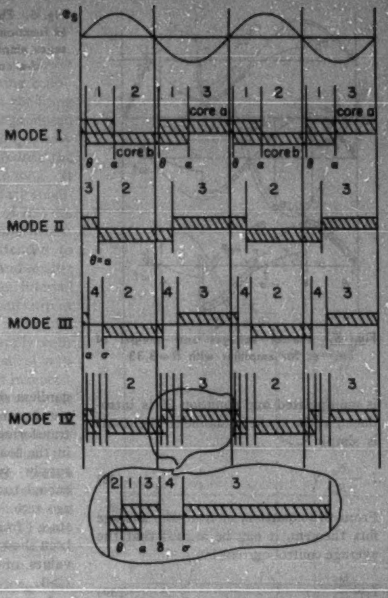


Fig. 4. Sequences of saturation states defining the four modes. Bars below and above axes indicate intervals in which core b and core a are unsaturated. Saturation states for intervals shown are numbered

Mode I

Suppose that core *b* is saturated. The magnetomotive force on core *b* is given by equations 4, 18, and 19.

$$F_b=N_g i_L-N_c i_c=\frac{2N_g}{R_L'}\left(e_s-\frac{E_c}{N}\right)=2N_g i_L \qquad (29)$$

Core *b* will return to its unsaturated condition when the magnetomotive force on the core becomes zero. Equation 29 shows that core *b* returns to its unsaturated condition when both the load current and control current become zero. This occurs when

$$e_s=\frac{E_c}{N} \qquad (30)$$

The supply voltage is negative during the half-cycle when core *b* is saturated. Equation 30 shows that core *b* remains saturated for a short interval after the beginning of the positive half-cycle of the supply voltage. The "cutoff angle" θ for core *b* is found from equation 30 as follows

$$\theta=\sin^{-1}\frac{E_c}{NE_s} \qquad (31)$$

The cutoff angle for core *a* one half-cycle later is found to be the same. The cutoff angles are shown in Fig. 4.

The firing angle for mode I can be found by taking the integral of equation 28 between adjacent cutoff angles

$$\int_\theta^{\pi+\theta}(e_a-e_b)d\omega t=0 \qquad (32)$$

As shown in Fig. 4 in the interval $\theta\le\omega t\le\alpha$, the circuit is in state 1. In the interval $\alpha\le\omega t\le\pi+\theta$ the circuit is in state 2. In the interval $\theta\le\omega t\le\alpha$, therefore, the voltage difference is

$$e_a-e_b=\frac{E_c}{N}$$

In the interval $\alpha\le\omega t\le\pi+\theta$, the voltage difference is

$$e_a-e_b=-e_s\left(1-\frac{R_L}{R_L'}\right)+\frac{E_c}{N}\frac{R_L}{R_L'}$$

The equivalent statement of equation 32 becomes

$$\int_\theta^\alpha\frac{E_c}{N}\,d\omega t-\int_\alpha^{\pi+\theta}\left[e_s\left(1-\frac{R_L}{R_L'}\right)-\frac{E_c}{N}\frac{R_L}{R_L'}\right]d\omega t=0 \qquad (32A)$$

The reactor voltages and the integrals in equation 32(A) are shown in Fig. 5 for a typical set of parameters.

In the interest of simplifying notation, two normalized parameters are introduced.

$$E\equiv\frac{E_c}{NE_s} \qquad (33)$$

$$R\equiv\frac{N^2R_L}{R_c} \qquad (34)$$

The integrations indicated in equation 32(A) are then performed, with the result

$$\cos\alpha=E[R\pi-\theta]-\cos\theta+E\alpha \qquad (35)$$

The cutoff angle θ has been shown in equation 31 to be a function of *E*. Equation 35 therefore gives the firing angle α as a function of *E* with *R* as a parameter. The solutions for α in this and in the three other modes are shown in Fig. 6.

The reactor voltage e_a and load current are plotted in Figs. 7 and 8 for two values of the resistance parameter *R*. The voltage e_a is similar to e_b, as shown in Fig. 5. For each value of the resistance parameter *R*, two firing angles are shown to illustrate the changes in the voltages and currents as the normalized input voltage is changed. The firing angle α is found from equation 35. The flux in core *b* is also shown in Fig. 8.

Since in mode I there is always at least one core unsaturated, the instantaneous values of control current and load current are always related through the turns ratio *N*. Thus, the rectified-average value of the load current is also related to the average control current by the turns ratio *N*.

$$I_L=NI_c \qquad (36)$$

The load current, from equation 29, can be formally averaged to obtain its rectified average value

$$I_L=\frac{1}{\pi}\int_\alpha^{\pi+\theta}\left(e_s+\frac{E_c}{N}\right)\frac{1}{R_L'}\,d\omega t \qquad (37)$$

This equation is integrated between the limits shown. If the resulting expression

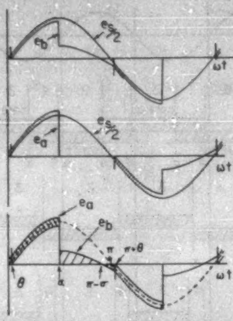
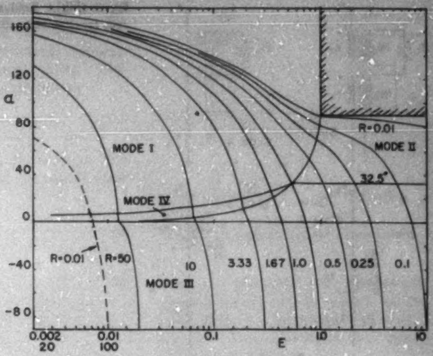


Fig. 5. Reactor voltages and integral of $e_a - e_b$ for amplifier with R = 3.33

Fig. 6. Firing angle as function of E for series amplifier with d-c control



is manipulated and equation 35 is introduced at the strategic moment, the result is simply

$$I_L = \frac{E_c}{R_c} N \tag{38}$$

From this equation, or from the average flux theorem, it can be argued that the average control current is

$$I_c = \frac{E_c}{R_c} \tag{39}$$

The voltage gain of the amplifier is therefore

$$G_v = N \frac{R_L}{R_c} \tag{40}$$

The results given by equations 36 through 40 are identical with those given by Storm for comparatively low control resistance.[1] This is expected since the same approximation to the magnetization curve is used.

It next becomes necessary to define the range of the parameters E and R for which mode I exists. In describing mode I, two important assumptions were made. The first was that at the cutoff angle θ in each half-cycle, the load current goes to zero sinusoidally. At $\omega t = \theta$ the load current becomes zero because the total generated voltage in the load circuit is zero. This was expressed in equation 30, which can be rewritten

$$\frac{E_c}{Ne_s} = 1 \tag{30A}$$

It follows from this that if E_c/N is larger than the peak value of supply voltage E_s

$$\frac{E_c}{NE_s} = E > 1 \tag{41}$$

and equation 30 cannot be satisfied regardless of the value of the resistance parameter R. In this situation, the transformed control voltage appearing in the load circuit is so large that the supply voltage can never be large enough to make the total load circuit voltage zero. Thus, the circuit cannot enter state 1 from either state 2 or 3, and it has been shown that mode I cannot exist for values of the control voltage so that $E > 1$.

A second important assumption was implied in writing the equations for mode I. The existence of state 1 in the sequence of saturation states implies that $\alpha > \theta$. If, however, the cutoff angle were to increase while the firing angle decreased until the two angles were equal, then state 1 would disappear from the sequence and the circuit would enter a second mode of operation. Equation 35 shows that an increase in cutoff angle and a decrease in firing angle accompany an increase of E, and that they will become equal at values of E depending upon the resistance parameter R. Thus a lower boundary on the firing angle for mode I can be found by setting α equal to the cutoff angle θ in equation 35.

$$2 \cos \alpha = ER\pi \tag{42}$$

It has already been shown that mode I cannot exist for values of $E > 1$. Solutions of equation 42 are therefore desired for values of $E \leq 1$, for which $E = \sin \theta = \sin \alpha$. Equation 42 then becomes

$$\tan \alpha = \frac{2}{R\pi} \tag{43}$$

It is seen immediately that the right side of equation 43 is always positive. The firing angle on this boundary curve is always found between zero and 90 degrees. The boundary, shown in Fig. 6, is simply a plot of

$$\alpha = \sin^{-1} E \tag{44}$$

It should be noted that this is a boundary below which mode I cannot exist. It has not been shown, however, to be the greatest lower boundary on mode I. An attempt to prove this formally, in fact, has led to another mode called mode IV, to be discussed later.

The boundary of mode I, just determined, is defined by a change in the sequence of saturation states within the cycle. A second mode is suggested which is similar to mode I except that θ is everywhere equal to α.

Mode II

Mode II is characterized by an alternation of states 2 and 3. One of the two cores is always in the unsaturated state. This implies that a linear relation exists between the rectified-average value of the load current and the average value of the control current. Mode II is therefore a linear control region. It is shown schematically in Fig. 4(B).

If mode II is to exist over a finite range of the parameters, it is necessary merely to write the fundamental equation for the steady state and to define the region for which it is valid. The voltage-difference integral is taken over the interval of state 2. The voltage e_a is zero in this interval. From equation 15 the integral is written

$$\int_{\alpha}^{\pi+\alpha} e_b d\omega t = \int_{\alpha}^{\pi+\alpha} \left[e_s\left(1 - \frac{R_L}{R_L'}\right) - \frac{E_c}{N} \frac{R_L}{R_L'} \right] d\omega t = 0 \tag{28-II}$$

The solution is

$$\cos \alpha = ER \frac{\pi}{2} \tag{45}$$

This is identical with equation 43, showing

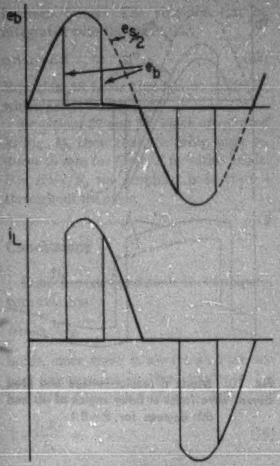


Fig. 7. Mode I reactor-voltage and load-current wave forms at firing angles of 60 and 120 degrees for R=50

that mode II includes the lower boundary defined for mode I.

In Fig. 6, equation 45 is used to plot the curves of firing angle as a function of E for mode II with R as the parameter. Figs. 9 and 10 show voltage and current wave forms for this mode for two values of the parameter R. On each diagram, firing angles of 40 to 60 degrees are shown.

In writing this integral equation it is tacitly assumed that at the angle α, the voltage e_b is positive. Reference to the polarities on the circuit of Fig. 1 shows that this is necessary if core b is to become unsaturated at the angle α when core a saturates. The limit for mode II will occur when e_b at the firing angle is exactly zero. This condition may be found by setting equation 15 equal to zero and evaluating at $\omega t=\alpha$. The result is

$$\sin\alpha=ER \tag{46}$$

Equation 46 may be divided by equation 45, yielding

$$\tan\alpha=\frac{2}{\pi} \tag{47}$$

$$\alpha=32.5^\circ \tag{48}$$

The interpretation of this equation is especially important since it suggests the form that mode III will take. On the boundary it was shown that the firing angle was given by equation 46. This can be rearranged slightly to read

$$\frac{E_s\sin\alpha}{R_L}=\frac{E_c}{R_c}N \tag{49}$$

Equation 49 shows that on the lower boundary of mode II, the limiting condition has been reached where at the firing angle of core a the load current produced by the supply voltage connected directly to the load is just enough to reduce the magnetomotive force on core b to zero. It has been shown that this boundary condition is reached at a firing angle of 32.5 degrees regardless of the circuit parameters.

This argument is easily extended to larger values of control voltage and smaller firing angles. There will be an interval after core a fires in which the load current and the control current flow independently. In this interval the supply voltage is small enough and the control voltage large enough so that the net magnetomotive force on core b at the firing angle α is not reduced to zero. The period in which the load and control currents flow independently will terminate when the supply voltage has become large enough to reduce the magnetomotive force on core b to zero. This situation is found only below the boundary defined by equation 46, however, and introduces a finite period of state 4 into the sequence. This new sequence defines a new mode, which is mode III.

Mode III

In mode III, there is an interval after the firing of core a in which both cores are saturated. This sequence is shown in Fig. 4(C). A positive voltage will appear on core b when the load current has attained the necessary value to make the magnetomotive force on the core equal to zero. The angle σ at which this occurs is defined by equation 46.

$$\sin\sigma=ER \tag{50}$$

The angle σ is plotted as a function of E with R as the parameter in Fig. 11. On the lower boundary of mode II, $\alpha=32.5$ degrees, it was shown that $\alpha=\sigma$. As expected, therefore, the curves of σ from Fig. 11 and the curves of α from Fig. 6 for $R\leq 1$ intersect on the boundary curve $\alpha=32.5$ degrees. For $R=1$, the curve of σ coincides with the boundary curve $\sin\alpha=E$, defined in equation 44.

Equation 28 is again applied to find α as a function of E. In mode II, the integral was taken between α and $\pi+\alpha$. In mode III the reactor voltage is zero between α and σ, so the integral is written

$$\int_\sigma^{\pi+\alpha}e_b\,d\omega t=\int_\sigma^{\pi+\alpha}\left[e_s\left(1-\frac{R_L}{R_L'}\right)-\frac{E_c}{N}\frac{R_L}{R_L'}\right]d\omega t=0 \tag{28-III}$$

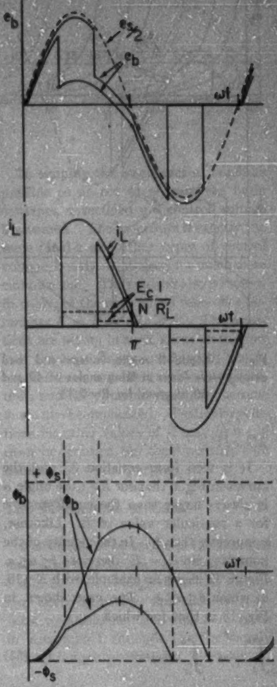


Fig. 8. Mode I reactor-voltage, flux, and load-current wave forms at firing angles of 60 and 120 degrees for R=3.33

The solution may be seen in the following equation

$$\cos\alpha-\alpha\sin\sigma=(\pi-\sigma)\sin\sigma-\cos\sigma \tag{51}$$

This equation is solved for the firing angle α, which is plotted as a function of E, with R as a parameter, in Fig. 6. It is seen from these figures that the firing angle in mode III extends to negative values. Typical voltage and current wave forms for mode III are shown in Fig. 12.

Between the firing angle defined in equation 51 and the angle σ defined by equation 50, it has been shown that an interval of state 4 exists in which both cores are saturated. It follows that mode III is not a linear control region. The general relation may be stated without further proof that for the load current to be a linear function of the control current, the circuit must be in a mode in which there is always at least one core unsaturated.

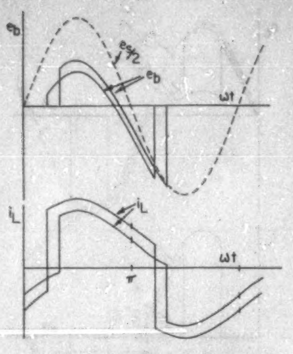


Fig. 9. Mode II reactor-voltage and load current wave forms at firing angles of 40 and 60 degrees for R=0.33

Mode IV

It is seen from equation 35 and the curves in Fig. 6 that for $R\geq 1$, the angle α is always larger than the cutoff angle θ for a particular value of E. Likewise, σ is greater than θ. In the vicinity of the boundary $\sin\alpha = E$, therefore $\theta\leq\alpha\leq\sigma$. Figure 13 shows an example, with $R=10$, in which $\theta\leq\alpha\leq\sigma$. The case shown in Fig. 13 is that for which

$$\int_{\theta}^{\alpha} e_b d\omega t+\int_{\alpha}^{\sigma} e_b d\omega t=0 \tag{52}$$

Core b is caused to return to the saturation point at the angle σ. In this example, the voltage e_b goes positive again at σ, and the flux change in the core reverses direction just in time to prevent core b from saturating again.

For smaller firing angles, core b reaches saturation again at the angle δ defined by the integral equation

$$\int_{\theta}^{\alpha} e_b d\omega t+\int_{\alpha}^{\delta} e_b d\omega t=0 \tag{53}$$

In the interval $\delta\leq\omega t\leq\sigma$ the circuit is in state 4. The sequence shown in Fig. 4(A) has been changed to that shown in Fig. 4(D).

The firing angle at which the transition from mode I to mode IV occurs can be determined by writing equation 52 in detail, using equation 8 and equation 15. When the integrations are performed, the result is found to be

$$\left(1-\frac{2R_L}{R_L'}\right)\cos\alpha+\left[2\left(1-\frac{R_L}{R_L'}\right)\sin\sigma-E\right]\alpha$$
$$=2\left(1-\frac{R_L}{R_L'}\right)(\cos\sigma+\sigma\sin\sigma)-E\theta-\cos\theta \tag{54}$$

This is a difficult equation to solve for α. Solutions were obtained directly by performing the graphical integrations indicated in Fig. 13 for various values of E. The resulting series of points defines the boundary between mode I and mode IV shown in Fig. 6.

In mode IV, the integral of the voltage difference e_a-e_b must still be zero in the steady state. The integrals are set up in a manner similar to those for the other three modes.

$$\int_{\theta}^{\alpha}\left(\frac{e_s}{2}+\frac{E_c}{2N}\right)d\omega t$$
$$=\int_{\sigma}^{\pi+\theta}\left[e_s\left(1-\frac{R_L}{R_L'}-\frac{E_c}{N}\frac{R_L}{R_L'}\right]d\omega t\right. \tag{28-IV}$$

The solution is

$$(\cos\theta-\cos\alpha)-E(\alpha-\theta)=2R[(\cos\sigma+\cos\theta)-\sin\sigma(\pi+\theta-\sigma)] \tag{55}$$

Solutions of equation 55 give the firing angle for mode IV as a function of E with R as a parameter, and are plotted in Fig. 6.

This reasoning holds for mode IV until the firing angle α reduces to the cutoff angle θ. In this case, the left side of equation 28-IV is zero, and the equation reduces to

$$\int_{\sigma}^{\pi+\alpha}\left[e_s\left(1-\frac{R_L}{R_L'}\right)-\frac{E_c}{N}\frac{R_L}{R_L'}\right]d\omega t=0 \tag{28-III}$$

Equation 54 is seen to be the defining equation, for mode III. This shows therefore that mode III and mode IV coincide on the boundary $\sin\alpha=E$.

In this discussion, four modes have been described. It has been shown that every part of Fig. 6 is included in one of the four modes except for the excluded region

$90°<\alpha<180°$
$E>1$

It was shown that if the circuit is in any one of the four modes, it cannot be brought into the excluded region through changes in the normalized voltage E or the normalized resistance R. It is therefore concluded that, subject to the assumed magnetic properties, all possible sequences of saturation states have been accounted for.

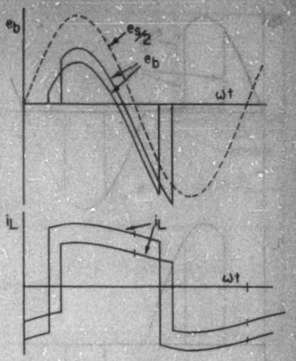


Fig. 10. Mode II reactor-voltage and load current wave forms at firing angles of 40 and 60 degrees for R=0.1

Series Amplifier with Rectified A-C Control

One of the factors which adds to the complication of the analysis of the series amplifier with d-c control is the fact that the direct voltage is at times larger than the a-c supply voltage. This causes the reactor voltage in states 2 and 3, equations 15 and 20, to go through zero at the angles σ and $\pi-\sigma$. That the control voltage is finite when the supply voltage goes through zero also accounts for the fact that the cutoff angle differs from zero. A considerable simplification in the analysis results if the control voltage is assumed to be a full-wave-rectified sine wave of the same frequency and phase as the supply voltage. The amplitude of the voltage is the variable. If the circuit is assumed to be in mode I, the voltage-difference integral can be written for the interval $0\leq\omega t\leq\pi$.

$$\int_{0}^{\alpha}\frac{E_c}{N}\sin\omega t\, d\omega t-$$
$$\int_{\alpha}^{\pi}\left[E_s\left(1-\frac{R_L}{R_L'}\right)-\frac{E_c}{N}\frac{R_L}{R_L'}\right]\times$$
$$\sin\omega t\, d\omega t=0 \tag{56}$$

The firing angle can be found by integrating equation 56.

$$\cos\alpha=\frac{E(1+2R)-1}{E+1} \tag{57}$$

It should be noted with care that when the symbol $E=E_c/NE_s$ is used, the quantity E_s is the peak value of the rectified sine wave used for control, not the average value.

Equation 57 is valid so long as the conditions for mode I are not violated. For $E>1$, the voltage applied to reactor b during the interval $0\leq\omega t\leq\alpha$ is negative. State 1 therefore drops out of the sequence, and the amplifier is in mode II. The first integrand of equation 56 becomes the same as the second, but with a plus sign between terms. The equation can be integrated with the result that

$$\cos\alpha = ER \qquad (58)$$

Equation 58 holds throughout mode II, which exists for all $E \geq 1$.

Equations 57 and 58, which are plotted in Fig. 14, show that the firing angle reduces to zero for $E = 1/R$ for either mode. For $E > 1/R$, the amplifier is in state 4 throughout the cycle.

Conclusions

In both mode I and mode II the ampere-turn relation

$$N_cI_c = N_gI_L$$

holds, since there is always at least one core unsaturated in these modes. If a normalized output voltage is defined as

$$V = \frac{I_LR_L}{E_s} \qquad (59)$$

the normalized gain can be written

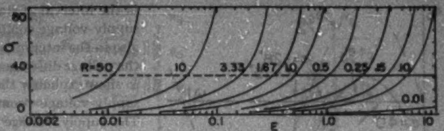
$$\frac{V}{E} = R \qquad (60)$$

where the normalized input voltage and the resistance parameter are defined by equation 33 and equation 34 respectively. Applied to the case of rectified a-c control, the control voltage E_c is replaced by the average value of the rectified a-c control voltage. Equation 60 becomes

$$\frac{V}{2E/\pi} = R \qquad (60A)$$

The results of this analysis may be of some use in applications where the series amplifier is still used. The important results of the analysis, however, is the method of approach. Two elementary and well-known theorems have been used to obtain equation 28. This equation has then been used to define the modes of operation and to find the firing angle as a function of control voltage. The roles of control-circuit impedance and control voltage in determining the mode of operation have been shown. For other assumed control voltages, magnetic characteristics, and perhaps such things as time-varying impedances, the integration indicated in equation 28 can be performed over the piecewise linear ranges corresponding to the saturation states in the modes. This has been done for hysteresis loops having slope and width, which complicates the analysis somewhat. The important point to be made, however, is that although it may be difficult to obtain particular solutions, the approach to the problem is unchanged. Often one can predict the effects of such disturbing phenomena without carrying out the detailed solutions.

Fig. 11. Curves of σ as function of E with R as parameter



No attempt has been made here to discuss all of the observed properties of the series amplifier. These are well known, and are covered thoroughly in the classic works of Storm.[1-3] In Part II it will be shown that equation 28 also gives a simple and clear picture of the mechanisms of gain and time constant of the series amplifier.

Part II—The Mechanisms of Gain and Time Constant

In Part I, the control-circuit voltage and resistance were included in the equations for the series amplifier. This was a crucial step, since it was only by this that it was possible to make effective use of the average-flux theorem and the symmetry theorem, from which was deduced the important steady-state equation

$$\int_{\omega t}^{\omega t+\pi}(e_b - e_a)d\lambda = 0 \qquad (28)$$

Careful examination of this equation in Part I showed that it could be used to define the four characteristic steady-state modes of operation of the amplifier with d-c control, and the two characteristic modes of operation of the amplifier with rectified a-c control. It showed also that altering the control voltage and the assumptions regarding the magnetization curve would alter the details of the solution, but not the method of approach so long as the control voltage is periodic at twice the power supply frequency.

Part II goes little further than to interpret equation 28. It is shown that this equation contains all the information necessary to give a clear picture of the mechanism of gain, and the mechanism of time constant, or delay, and their relation to one another. It is shown how the control voltage physically controls the amplifier and that the gain and time constant are consequences of the same basic mechanism of control. The results as applied to the series amplifier are incidental. It is hoped that the concepts developed will find application in the analysis of magnetic amplifiers in general.

To simplify the mechanics as much as possible so as not to obscure the basic concepts, a rectified a-c control voltage is assumed. The procedure is exactly the same for d-c and other types of control voltage, but the details become much more cumbersome. The analysis also is limited to values of the normalized resistance parameter $R = R_LN^2/R_c \gg 1$. The equations are set up in such a way as to show how the argument proceeds when this is not the case, but the graphic interpretations and the dynamic equations become much more complicated. Particular solutions for other values of R may in many cases be required, but their inclusion here does not serve the purpose.

The Mechanism of Gain

The circuit to be used and the assumed magnetic characteristics are shown in Figs. 1 and 2. The voltage equations for the two electrical meshes and the two magnetic equations for the cores were given in equations 1 through 4. The supply voltage is again given by

$$e_s = E_s \sin\omega t$$

The d-c control voltage in the equations must be replaced by

$$e_c = E_c \sin\beta \qquad (61)$$

where the argument β is measured in radians from the beginning of the half-cycle in question and ranges over the values $0 \leq \beta \leq \pi$. The control voltage is a rectified sine wave of the same frequency and phase as the supply frequency. The amplitude is the variable. The positive polarity of the control voltage is taken the same as the positive d-c control voltage.

In Part I the solutions for reactor voltages and currents in state 1 were given in equations 6 through 8. In state 2 the solutions were given by equations 13 through 15. In state 3 the solutions were given by equations 18 through 20. In state 4, the solutions were given by equations 21 through 24. These equations may be found by reference to Part I or they can easily be derived by inspection of the circuit.

It is convenient to take the integral of equation 28 over the half-cycle $0 \leq \omega t \leq \pi$.

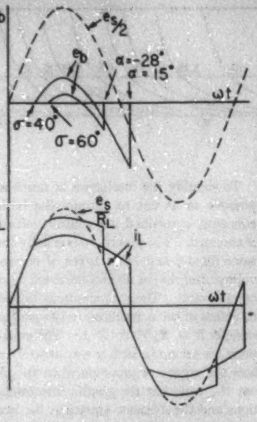


Fig. 12. Mode III reactor-voltage and load-current wave forms at firing angles of 15 and −28 degrees, σ of 40 and 60 degrees

$$\int_0^\pi (e_b-e_a)d\omega t=0 \qquad (62)$$

In the interval $0\leq\omega t\leq\alpha$ the circuit is in state 1. Neither core is saturated and the voltage difference is determined from equations 7 and 8.

$$e_b-e_a=-e_c/N \quad (0\leq\omega t\leq\alpha) \qquad (63)$$

During the remainder of the half-cycle, core a is saturated and the voltage e_a is zero. The voltage difference is just e_b, and is given by equation 15.

$$e_b-e_a=e_s\left(1-\frac{R_L}{R_L'}\right)-\frac{e_c}{N}\frac{R_L}{R_L'}(\alpha\leq\omega t\leq\pi) \qquad (64)$$

Equation 62 then becomes, for the steady state

$$-\int_0^\alpha \frac{e_c}{N}d\omega t+\int_\alpha^\pi\left[e_s\left(1-\frac{R_L}{R_L'}\right)-\frac{e_c}{N}\frac{R_L}{R_L'}\right]d\omega t=0 \qquad (65)$$

The half-cycle chosen is one in which the supply voltage is positive. In the negative half-cycle, however, e_b-e_a is found to be the same, i.e., it is periodic at twice the supply frequency.

The voltage wave forms involved are shown in Fig. 15. The equivalent circuits for the two saturation states are also shown. The voltage wave form consists of a control-voltage component $-e_c/N$ in the interval of state 1, $0\leq\omega t\leq\alpha$, a control voltage component

$$-\frac{e_c}{N}\frac{R_L}{R_L'}$$

in the interval of state 2, $\alpha\leq\omega t\leq\pi$, and a supply-voltage component $e_s(1-R_L/R_L')$ also in the interval of state 2. In Fig. 16 the voltage difference in Fig. 15 is redrawn to show explicitly the control-voltage and supply-voltage components of e_a-e_b. The supply voltage and half the supply voltage are also included in Fig. 16. The particular example chosen for Figs. 15 and 16 is the one that was used for Fig. 17 with $R=3.33$. The firing angle is 120 degrees. This example is used because the voltage difference is of proportions which can easily be seen when drawn to the same scale as the supply voltage.

Virtually all of the generalizations regarding the control and response of the magnetic amplifier will be based on the diagram in Fig. 16 and the voltage-difference integral, equation 62.

The equivalent circuits in Fig. 15 show the physical source of the difference-voltage components. In state 1, the difference between the two reactor voltages is simply the control voltage reduced by the turns ratio N, since no current flows in the control winding. In state 2, core a is saturated and core b is unsaturated. Core b acts as a perfect transformer during this period. The control-voltage and supply-voltage components of e_b can be found separately by considering one voltage generator at a time. Equation 64 was obtained in this manner. If equation 64 is rewritten in terms of the parameter R, the voltage difference is separated into supply-voltage and control-voltage components as follows

$$e_b-e_a=e_s\left(\frac{1}{1+R}\right)-\frac{e_c}{N}\frac{R}{1+R}\quad \alpha\leq\omega t\leq\pi \qquad (66)$$

It will be shown that when the supply-voltage component is small and the control-voltage component large, the gain of the amplifier is large. From equation 66 this is found to be the case for large values of R. In Part I, in fact, R was found in equation 60 to be equal to the normalized voltage gain of the amplifier. The only limit to the gain of the amplifier is the existence of the supply-voltage component in equation 66. This exists because of the fact that load current is always required to flow through an unsaturated core in mode I.

The wave forms in Fig. 16 represent the steady state for a particular value of the control voltage e_c/N. The integrals of the components that make up e_b-e_a must sum to zero. If the control voltage is increased by an amount $\Delta e_c/N$, the control voltage components in Fig. 16 increase in the negative direction by the amount $\Delta e_c/N$ from 0 to α, and

$$\frac{\Delta e_c}{N}\frac{R_L}{R_L'}$$

from α to π. The positive component of the voltage difference is derived from the load current and is a section of the sine wave, $e_s(1-R_L/R_L')$, whose magnitude is independent of the control voltage. The additional positive voltage integral required to balance the added negative integral is therefore obtained by a decrease in the firing angle α. Associated with this change of firing angle is a corresponding change in the average output voltage E_0. In Fig. 16, an increment of control voltage is shown producing an increase in the negative component of the integral of e_b-e_a. The area shaded under the positive voltage component $e_s(1-R_L/R_L')$ is therefore equal to the negative integral component. The change in firing angle shown is that required to produce this equality. The large unshaded area under the voltage e_s is very nearly the increase of the output voltage integral.

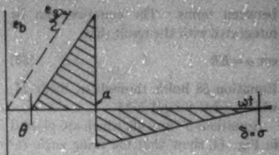


Fig. 13. Enlarged view of reactor voltage e_b on the boundary between mode I and mode IV for R=10

Before the increment of control voltage is applied, the control voltage components of e_b-e_a produce a negative voltage integral, as shown in Fig. 16. This quantity will be identified by the following symbol

$$\int e-=\frac{E_c}{N}(1-\cos\alpha)+\frac{E_c}{N}\frac{R_L}{R_L'}(1+\cos\alpha)$$

In the remainder of the argument it will be assumed that $R>>1$. Then

$$\int e-=\int_0^\pi\frac{e_c}{N}d\omega t=2\frac{E_c}{N} \qquad (67)$$

$$\int e+=\int_\alpha^\pi e_s\left(1-\frac{R_L}{R_L'}\right)d\omega t$$

The supply-voltage component of e_b-e_a is identified by

$$\int e+=\int_\alpha^\pi e_s\left(\frac{1}{1+R}\right)d\omega t \qquad (68)$$

When an increment of control voltage is applied, the firing angle begins to de-

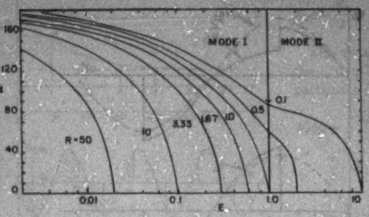


Fig. 14 (left). Firing angle as function of E for series amplifier controlled with full-wave-rectified sine wave of voltage

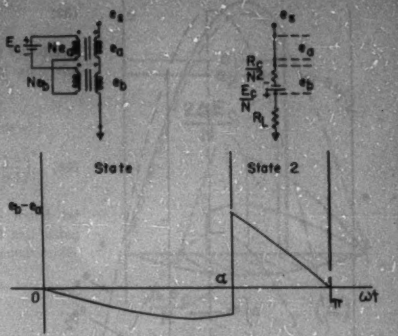


Fig. 15 (right). Diagram of voltage difference e_b-e_a for one half-cycle

crease, and continues to do so until the new steady-state condition shown in Fig. 16, is reached. When the new steady-state condition has been reached, the firing angle $\alpha'=\alpha+\Delta\alpha$ will be reduced sufficiently so that

$$\Delta\int e_- = \Delta\int e_+$$

Thus

$$\frac{2\Delta E_c}{N} = \int_{\alpha'}^{\alpha} e_s\left(\frac{1}{1+R}\right) d\omega t \qquad (69)$$

This voltage integral can on the right of equation 69 be treated as the increment of control-voltage integral. For larger values of R, the amplitude of the sine wave

$$e_s\left(\frac{1}{1+R}\right)$$

is smaller. Therefore, to satisfy equation 69 the change of firing angle will be increased. But the output is governed directly by this firing angle as shown in Fig. 16. The change in output voltage integral is approximately given by

$$\Delta E_o\pi = \int_{\alpha'}^{\alpha}\left[e_s - e_s\left(\frac{1}{1+R}\right)\right] d\omega t \approx \int_{\alpha'}^{\alpha} e_s\left(\frac{R}{1+R}\right) d\omega t \qquad (70)$$

The ratio of this to the equivalent input voltage integral from equation 69 is equal to the ratio of the increment in average output voltage to the increment in average input voltage.

$$\frac{\Delta E_o}{\Delta 2E_c/N\pi} = R \qquad (71)$$

This expression for the voltage gain is seen to be equivalent to the result obtained in equation 60.

There are two important points to be made regarding the expression for normalized voltage gain of the amplifier. The first is that by means of the ampere-turn equality the voltage gain in equation 60 was written with no restrictions on the value of R. The second point is that from the wave form of the voltage difference in Fig. 16, the voltage gain has been read directly. This step is of considerable significance because a simple diagram has been found which shows the mechanism by which a change in control voltage produces a change in output voltage. By extending these concepts a bit further, it is not difficult to see that the time constant is explained in terms of this same mechanism.

The Mechanism of Time Constant

The diagram in Fig. 16 and the arguments that were used to develop the mechanism of gain also imply the mechanism of time constant. Before enlarging on this point, however, the notion of "flux charge" will be developed.

In Fig. 17 flux wave forms were shown corresponding to the reactor voltages in the steady state. It was noted that for core b at $\omega t=0$, the flux starts from saturation and increases to a maximum value at $\omega t=\pi$. This maximum flux increment is called the flux charge. It corresponds to an integral of voltage over a finite angular interval, and therefore it has no meaning as a continuous function of the angular variable ωt. For physical reasons soon to be apparent the value is defined of the function flux charge at the end of the half-cycle in which it is accumulated. It is a discrete function of the index n, which will be used to number the half-cycles of the supply voltage. Note that the flux charge gained by core b on the first half-cycle is discharged on the second half-cycle. At the end of the nth half-cycle the flux discharged in the core during the half-cycle will be represented and called "flux charge," recognizing that the voltage integral that produced it occurred on the previous half-cycle. Further, voltage integrals will be called "fluxes" although this is not quite correct dimensionally.

To rid the equations of the constant ω, the normalized flux will be used

$$\phi = BA\omega$$

where B is flux density and A is the cross-sectional area of the magnetic path.

The discrete function flux charge therefore has a value at the end of each half-cycle equal to the flux discharged by the firing core in that half-cycle. Its value is measured alternately in cores a and b. It can easily be shown that the flux discharged by the firing core in a given half-cycle and the output voltage integral during that half-cycle add up to a constant. The flux discharged by core a, for instance, in a particular half-cycle is

$$N_g\phi_a = \frac{1}{2}\int_0^{\alpha}\left(e_s+\frac{e_c}{N}\right) d\omega t$$

The integral of half the output voltage in that half-cycle is

$$\frac{1}{2}E_o\pi = \frac{1}{2}\int_{\alpha}^{\pi}\left(e_s+\frac{e_c}{N}\right) d\omega t$$

It is seen that the sum of the two equals one-half the integral of the supply voltage plus the control voltage. Fig. 18 shows the relation of these to the index n.

In the steady state, equation 62 implies that the flux discharged by one core is equal to the flux charge gained in the other during any given half-cycle. If a half-cycle is considered in which core a fires, the flux discharged by core a is already determined. The integral of the voltage difference gives the increment of flux charge during the half-cycle, and appears directly as an increment in the flux charge gained by core b during that half-cycle. Thus, in the steady state

$$\delta N_g\phi = \int_0^{\pi}(e_b-e_a) d\omega t = 0 \qquad (62A)$$

The right side of this equation is expanded in equation 65.

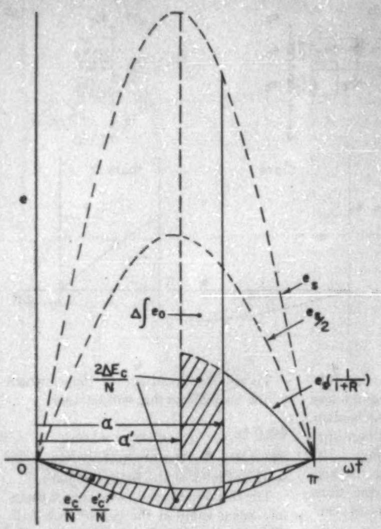


Fig. 16 (left). Diagram of voltage difference $e_b - e_a$ in Fig. 15 redrawn to show negative and positive components

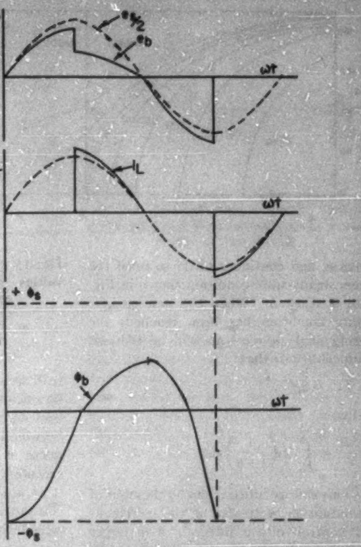


Fig. 17 (right). Reactor voltage, load-current, and flux wave forms for rectified sine wave control, R=3.33

An increment of control voltage is now applied. If it is assumed that $R>>1$ and that during the first half-cycle the firing angle does not change, equation 65 yields the following

$$\delta N_g\phi_1 = \int_0^\pi (e_b - e_a)d\omega t \approx \int_0^\pi \frac{\Delta e_c}{N}\, d\omega t = -\frac{2E_c}{N} \quad (72)$$

This is the amount by which the flux discharged by core *b* decreases in the second half-cycle. A diagram showing these relations is shown in Fig. 19(A). In this diagram, the voltage that discharges core *b* on the second half-cycle is shown superimposed on the voltage that discharges core *a* on the first half-cycle. The firing angle is shown decreased to account for the decrease in the flux charge gained by core *b* on the first half-cycle, given by equation 72. It is not strictly true that the firing angle does not change on the first half-cycle, because the voltage discharging core *a* has increased by a small amount. This correction factor will be included later in the argument. At present, the important point to be made is that the integral of the voltage difference gives the change in flux charge on any half-cycle, and that this increment can be readily identified on a diagram of the voltage difference such as Fig. 16.

Consider now the events that occur in the general or *n*th half-cycle. The firing angle is α_n. The new control voltage is $e_c' = e_c + \Delta e_c$. The increment of flux charge on the *n*th half-cycle is then

$$\delta N_g\phi_n = N_g(\phi_{n+1} - \phi_n) = \int_{(n-1)\pi}^{n\pi} (e_b - e_a)d\omega t \quad (73)$$

This can be expanded as follows

$$\delta N_g\phi_n = -\int_0^{\alpha_n} \frac{e_c'}{N}\, d\omega t + \int_{\alpha_n}^{\pi} \left[e_s\left(\frac{1}{1+R}\right) - \frac{e_c'}{N}\frac{R_L}{R_L'} \right] d\omega t \quad (74)$$

If it is assumed that $R>>1$, equation 74 can be rewritten

$$\delta N_g\phi_n = -\int_0^\pi \frac{e_c'}{N} d\omega t + \int_{\alpha_n}^{\pi} e_s\left(\frac{1}{1+R}\right) d\omega t \quad (75)$$

The integration in equation 75 is then performed.

$$\delta N_g\phi_n = -\int_0^\pi \frac{e_c}{N} d\omega t + \int_{\alpha}^{\pi} e_s\left(\frac{1}{1+R}\right) d\omega t - \int_0^\pi \frac{\Delta e_c}{N}\, d\omega t + \int_{\alpha_n}^{\alpha} e_s\left(\frac{1}{1+R}\right) d\omega t \quad (76)$$

The first half of equation 76 is zero from equation 65. Equation 76 then becomes

$$\delta N_g\phi_n = -\frac{2\Delta E_c}{N} + E_s\left(\frac{1}{1+R}\right)(\cos\alpha_n - \cos\alpha) \quad (77)$$

The term cos α was given by equation 57

$$\cos\alpha = \frac{E(1+2R)-1}{E+1} \quad (57)$$

This is a function only of the normalized resistance and voltage parameters, and is not a function of the index *n*. It therefore can be treated as a constant.

The term cos α_n can be found by integrating the following in terms of the flux discharged on the *n*th half-cycle

$$N_g\phi_n = \int_0^{\alpha_n} \left(\frac{E_s}{2} + \frac{E_c'}{2N}\right) \sin\omega t\, d\omega t \quad (78)$$

$$N_g\phi_n = \frac{E_s}{2}(1+E')(1-\cos\alpha_n)$$

for $R>>1$, it is shown from equation 58 that $E'<<1$.

$$\cos\alpha_n = 1 - \frac{2N_g\phi_n}{E_s} \quad (79)$$

Equation 77 can then be rewritten

$$N_g\phi_{n+1} - N_g\phi_n = -\frac{2\Delta E_c}{N} + E_s\left(\frac{1}{1+R}\right)\left(1 - \frac{2N_g\phi_n}{E_s} - \cos\alpha\right) \quad (80)$$

$$N_g\phi_{n+1} - \left(1 - \frac{2}{1+R}\right)N_g\phi_n = -\frac{2\Delta E_c}{N} + E_s\left(\frac{1}{1+R}\right)(1-\cos\alpha) \quad (81)$$

Equation 81 is a first-order difference equation. The right side of the equation is a constant driving function. The complementary solution is exponential, and since the driving function is a step function multiplied by a constant, the total re-

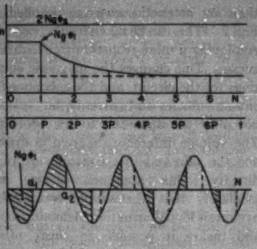


Fig. 18. The discrete function "flux charge" as function of index n. Index n placed at end of half-cycle to which it refers

sponse, which is a discrete function of n, has an exponential envelope. The complementary equation is then

$$\left[D-\left(\frac{R-1}{R+1}\right)\right]N_g\phi_n=0 \tag{82}$$

The solution is assumed

$$N_g\phi(t)=C\epsilon^{-nP/T} \tag{83}$$

The constant T is the time constant. The constant P is the time of one half-cycle, as shown in Fig. 18. Equation 83 is the flux charge as a function of time. The variable in the exponent nP indicates that the function is to be evaluated only at these discrete values of time. It can be written as a function of the index n to conform with the notation in equation 81 by simply regrouping.

$$N_g\phi_n=C[\epsilon^{-P/T}]^n=C[\epsilon^m]^n \tag{84}$$

where $m=-P/T$. Equation 84 is recognized to be in the form usually selected for the complementary function of the first-order difference equation. The symbol C is an arbitrary constant; m is a parameter. By placing this form of $N_g\phi_n$ in equation 82, the solution can be found in the usual manner. The result is

$$\epsilon^m=\left(\frac{R-1}{R+1}\right) \tag{85}$$

After the new steady state is reached, equation 81 can be written since it is known that $N_g\phi_{n+1}=N_g\phi_n=N_g\phi'$

$$N_g\phi'\left(1-\frac{R-1}{R+1}\right)=-\frac{2\Delta E_c}{N}+E_s\left(\frac{1}{1+R}\right)(1-\cos\alpha) \tag{86}$$

$$N_g\phi'=-\frac{\Delta E_c(1+R)}{N}+\frac{E_s}{2}(1-\cos\alpha) \tag{87}$$

The initial condition is found in the same manner. Before the control voltage is increased it has the value e_c. Also, $N_g\phi_{n+1}=N_g\phi_n=N_g\phi$.

$$N_g\phi=\frac{E_s}{2}(1-\cos\alpha) \tag{88}$$

The total solution is found by adding the complementary solution and the particular integral

$$N_g\phi_n=-\frac{\Delta E_c(1+R)}{N}+\frac{E_s}{2}(1-\cos\alpha)+c\left(\frac{R-1}{R+1}\right)^n \tag{89}$$

The arbitrary constant is found by putting $n=0$ in equation 89 and using the initial value of $N_g\phi_n$ from equation 88.

$$C=\frac{\Delta E_c(1+R)}{N} \tag{90}$$

Equation 89 then becomes

$$N_g\phi_n=\frac{E_r}{2}(1-\cos\alpha)-\frac{\Delta E_c(1+R)}{N}\left[1-\left(\frac{R-1}{R+1}\right)^n\right] \tag{91}$$

If it is recognized that the assumption of $R>>1$, already made, results in a response which is slow compared to the time of a half-cycle P, the exponent in equation 84 is small compared to unity. In that event, the exponential can be fairly well represented by a few terms of its power series

$$\epsilon^{-P/T}=1+\left(-\frac{P}{T}\right)+\frac{(-P/T)^2}{2!}+\frac{(-P/T)^3}{3!}+\dots$$

If the first two terms only are taken, equation 84 can be equated to equation 85.

$$1+\left(-\frac{P}{T}\right)=\frac{R-1}{R+1}$$

$$\frac{P}{T}=2\left(\frac{1}{1+R}\right) \tag{92}$$

But $1+R\doteq R$ and $P=1/2f$. Thus the time constant becomes

$$T=\frac{1}{4f}R \tag{93}$$

This is seen to be the expression for time constant developed by Storm[8] for the series amplifier with comparatively low control resistance.

Fig. 19(B) shows the graphical representation of the change in flux charge in the nth half-cycle. These relations were written in terms of two adjacent half-cycles. The flux discharged in the first half-cycle, as shown in Fig. 18, is not affected by the increment of control voltage during that half-cycle. The exponential decay of flux charge, corresponding to an exponential rise in output voltage, therefore begins one half-cycle after the switch is closed, applying the increment of control voltage. The response of the amplifier can therefore be represented in terms of a d-c gain, one time constant, and a real displacement of one half-cycle. The real displacement of one half-cycle, however, is only approximate. The flux discharged does not change in the first half-cycle, but the firing angle, and therefore the output voltage, does change somewhat because the increment of control voltage discharges the firing core in a shorter period. If the quiescent firing angle is nearly 180 degrees, the change in firing angle will be as much as half that indicated for the first half-cycle in Fig. 19(A). The real displacement therefore may be as small as 1/4 cycle. Also, the displacement will depend upon the exact time at which the increment of control voltage is applied.

The voltage transfer function for the amplifier can therefore be written, assuming a half-cycle displacement

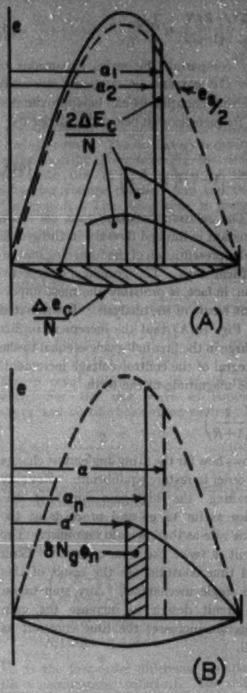


Fig. 19. Voltage diagrams which relate gain and time constant

$$\frac{E_o}{E_{in}}=\frac{R/N}{(1+Ts)}\epsilon^{-(\pi/\omega)s} \tag{94}$$

By comparing the series expansions of $\epsilon^{-(\pi/\omega)s}$ and $(1+\pi/\omega s)^{-1}$, one can easily see that equation 94 can be approximated at frequencies well below the carrier by

$$\frac{E_o}{E_{in}}=\frac{R/N}{(1+Ts)\left(1+\frac{\pi}{\omega}s\right)} \tag{94A}$$

The relation between gain and time constant is implied directly in the graphical representation of the foregoing mathematical steps. This graphical interpretation, in fact, is probably the most important result of the analysis. It was noted in Fig. 19(A) that the increment in flux charge in the first half-cycle is equal to the integral of the control-voltage increment. The magnitude of the term

$$e_s\left(\frac{1}{1+R}\right)$$

shows how far the firing angle must change in order to restore equilibrium. Roughly speaking, the high-gain amplifier has twice as far to go and proceeds at the same rate as the low-gain amplifier. The result is twice the time constant. Gain and time constant are the result of this one single mechanism. Any step taken in circuit design to increase the gain therefore increases the time constant as well.

Conclusions

All magnetic amplifiers that display the proper symmetry conditions and assume the magnetic properties of Fig. 2 yield diagrams of the voltage difference similar to that of Fig. 16. If the diagram for the parallel amplifier is drawn, it is easily seen that it has twice the gain and four times the time constant of the series amplifier. If external feedback is applied, the effect is merely to reduce the current flowing in the control winding by causing some of it to flow in the feedback windings. This reduces the voltage

$$e_s\left(\frac{1}{1+R}\right)$$

and increases the gain. Obviously, the time constant is also increased because the model which has been developed is unchanged.

In the case of the simple series and parallel amplifiers, the gain is relatively low and the control current is relatively large. The error in assuming the magnetic characteristic of Fig. 2 is small if the coercive force of the magnetic material is small and the hysteresis loop is nearly rectangular. When large amounts of positive feedback are used, or the current in the unsaturated reactor is blocked by means of rectifiers, the control current becomes of the same order of magnitude as the error assumed in making the assumptions on magnetic materials implied in Fig. 2. This can be taken into account by employing a more realistic magnetization curve. This changes the magnetic equations for the circuit, and the solutions for circuit voltages and currents, but it does not alter the approach to the problem. The voltage-difference diagram is drawn and the gain and time constant are read directly. This approach shows immediately that the simple rectangular hysteresis loop is an untenable assumption, and makes it possible to check other simple magnetic models fairly quickly. A subsequent paper will show the results of this procedure applied to the common "self-saturating" magnetic amplifiers.

References

1. SERIES-CONNECTED SATURABLE REACTOR WITH CONTROL SOURCE OF COMPARATIVELY LOW IMPEDANCE, H. F. Storm. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 756–65.

2. SERIES-CONNECTED SATURABLE REACTOR WITH CONTROL SOURCE OF COMPARATIVELY HIGH IMPEDANCE, *Ibid.*, pp. 1299–1309.

3. MAGNETIC AMPLIFIERS (book), *Ibid.*, John Wiley & Sons, Inc., New York, N. Y., 1953.

4. CONFERENCE ON MAGNETIC AMPLIFIERS. *AIEE Technical Publication T86.* "Recent Advances in the Theory of Magnetic Amplifiers," R. C. Barker. Joint AIEE, IRE and ISA., Institute of Radio Engineers, New York, N. Y., Instrument Society of New York, New York, N. Y., July 1956.

5. SERIES-CONNECTED SATURABLE REACTOR WITH CONTROL SOURCE OF COMPARATIVELY LOW IMPEDANCE, H. F. Storm. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 756–65.

◆

Discussion

H. C. Bourne, Jr. (University of California, Berkeley, Calif.): Part I of the paper is interesting as a good exercise in the steady-state solution of a piecewise linear circuit. As could be anticipated with these particular excitations, the transcendental equations encountered limit the usefulness of the results in any attempt to represent this magnetic amplifier as a systems component.

The discusser is frankly puzzled by the claims made and the results obtained in Part II of this paper. The author should certainly be aware of the paper by Johannessen.[1] Equation 28 is only a restricted form of the determining equation represented by equation 10 of the reference. With sinusoidal excitation and rectified a-c control the assumption $R>>1$ seems hardly necessary and, in fact, obscures important gain-bandwidth concepts. The mechanism of gain and time constant could not be expressed more clearly and in a more useful form than in equation 13 of Johannessen's paper, in which no assumption involving R is used. The dynamic transfer function of equation 15 of the reference with an appropriate change of symbols

$$G(s)=\frac{\frac{1}{N}\frac{R}{R+1}\left(1+\epsilon^{-\frac{\pi}{\omega}s}\right)}{1-\frac{R-1}{R+1}\epsilon^{-\frac{\pi}{\omega}s}}=\frac{K_1(z+1)}{z-K_2}$$

in which

$$z=\epsilon^{-\frac{\pi}{\omega}s}$$

is reduced to the author's result as a special case. The addition of a half-cycle delay to equation 94 hardly seems justified since the approximation involving the use of T for the time constant has already implied that this delay is negligible. There are other subtleties of this nature concerning the approximations involved in a simplification of $G(s)$, i.e.

$$\left(\epsilon^{-\frac{\pi}{\omega}s}\cong 1-\frac{\pi}{\omega}s \text{ and } \epsilon^{-\frac{\pi}{\omega}s}\cong 1+\frac{\pi}{\omega s}\right)$$

which call for some care in interpretation.

Neither the author nor Johannessen considers the case of dynamic operation in mode II. Dynamically, the behavior of the circuit is different in the two modes, as indicated by Ringelman and Fenaroli[2] and emphasized in the discussion of their paper. From gain-bandwidth considerations, the circuit as a systems component would seldom be designed to operate in this mode.

REFERENCES

1. ANALYSIS OF MAGNETIC AMPLIFIERS BY THE USE OF DIFFERENCE EQUATIONS, P. R. Johannessen. *Ibid.*, vol. 73, pt. I, 1954 (Jan. 1955 section), pp. 700–11.

2. THE SIMPLE REACTOR CIRCUIT; ITS OPERATION AND MODE TRANSITION, J. F. Ringelman, A. L. Fenaroli. *AIEE Transactions*, vol. 75, pt. I, 1956 (Jan. 1957 section), pp. 660–69.

R. C. Barker: The transcendental equations of Part I are those required to give the firing angle as a function of the control voltage and the circuit parameters. The appearance of these equations does not alter the fact that the static gain of the amplifier, as shown in equation 38 and in the conclusion, is the same constant as it is for rectified a-c control. The procedure for finding the frequency response is the same as for rectified a-c control, and the results are substantially the same. But the algebra is somewhat more complicated.

The author is grateful to Mr. Bourne for reminding us of the excellent work done at Massachusetts Institute of Technology relevant to this subject. The results obtained by Johannessen are virtually the same as those given in the foregoing, although his paper was written with a somewhat different purpose. The "determining equation" of the reference is the same as the author's equation 25 in Part I. Equation 25 was here written for the steady state. In Part II, the quantity "flux charge" is discussed at length, and it is argued that equation 25 holds in the dynamic case as well, leading to the interpretation of the voltage-difference integral as a flux-charge increment in equation 62(A). This simple physical identification of the induced voltage at the control terminals with the reactor fluxes is the reason for the author's, dealing with reactor fluxes directly rather than with output voltages. One can relate the output-voltage integral to the reactor fluxes, as pointed out prior to Equation 62(A), by writing

$$E_{o_n\pi}=\frac{2R}{1+R}[E_s(1+E_n)-N_g\varphi_n] \qquad (95)$$

This can be written by inspection from Fig. 16 as well. In this manner, the increment in output-voltage integral is related to the increment in flux charge by

$$\delta E_{o_n\pi}=\frac{2R}{1+R}[E_s(E_{n+1}-E_n)-\delta N_g\varphi_n] \qquad (96)$$

The assumption of $R>>1$ is made entirely to simplify the algebra in the example chosen for this paper, and does not alter the resulting gain or bandwidth. The procedure for writing the difference equation exactly is the same as that followed in the paper. The author recommends, as an exercise, starting with equation 67 and working the problem with no assumptions regarding the value of R. The algebra multiplies rapidly, but the conclusions are only slightly different. The increment in output-voltage integral, equation 70, becomes

$$\Delta E_{o\pi}=\int_{\alpha^1}^{\alpha}\left(e_s+\frac{e_c'}{N}\right)\left(\frac{R}{1+R}\right)d\omega t+\int_{\alpha}^{\pi}\frac{\Delta e_c}{N}\left(\frac{R}{1+R}\right)d\omega t \qquad (70\text{A})$$

Integration of equation 74 exactly gives in place of equation 77

$$\delta N_g\varphi_n=\frac{\Delta E_c}{N}\left[\cos\alpha_n-\left(\frac{R}{1+R}\right)\cos\alpha-\left(\frac{1+2R}{1+R}\right)\right]+\left(\frac{E_c-\Delta E_cR}{N(1+R)}+\frac{E_s}{1+R}\right)[\cos\alpha_n-\cos\alpha] \qquad (77\text{A})$$

The resulting difference equation is

$$N_g\varphi_{n+1}-\left(\frac{R-1}{R+1}\right)N_g\varphi_n=-\frac{2\Delta E_c}{N}\left(\frac{R}{1+R}\right)+\frac{E_s+E_c/N}{1+R}(1-\cos\alpha) \qquad (81\text{A})$$

This is slightly different from equation 81, and is equivalent to Johannessen's equation 10. The solution to equation 81(A) is

$$N_g\varphi_n=1/2(E_s+E_c/N)(1-\cos\alpha)-\frac{\Delta E_cR}{1+R}\left[1-\left(\frac{R-1}{R+1}\right)^n\right] \qquad (91\text{A})$$

Note that the same characteristic equation is obtained from all of these different equations. It is possible to find the roots of the first-order difference equation without applying Nyquist's criterion.

The author thoroughly agrees that the mechanisms of gain and time constant are clearly implied in the difference equation for the amplifier. The principal objective of Part II was, in fact, to enlarge upon the difference equation, giving physical and graphical interpretation to its terms. The author solved the difference equation for a step function applied at the input. The resulting time constant and the static gain are used to construct the transfer function. Johannessen took the Laplace transform of the difference equation to obtain his equation 15. This is the more accurate expression for the amplifier transfer function. If the exponentials are replaced by means of the approximate equality, for signal frequencies much lower than that of the carrier

$$\epsilon^{-\frac{\pi}{\omega}s}\doteq\frac{1}{1+\frac{\pi}{\omega}S}$$

Johannessen's equation 26 results directly

$$G(S)=\frac{R/N\left(1+\frac{\pi}{2\omega}S\right)}{1+\frac{\pi}{2\omega}(1+R)S}$$

The assumption made that signal frequencies are much lower than the carrier states that the numerator time constant is insignificant in the range for which the expression is valid. For any value of R, the derivation of the author's equation 92 requires the same assumption in order to represent the exponential by the first two terms of the Taylor series. Ignoring the author's delay term and retaining the $1+R$ from equation 92, one finds that the two results are the same and are based on the same approximations. Both yield the same gain-bandwidth product. At frequencies where the delay is important, or where the approximation by the transfer function of a continuous-data system becomes inadequate, other factors become important. The output is defined as an impulse function appearing at the end of the nth half-cycle. Since we are interested in the system viewpoint, it will be necessary in such cases to determine whether the load will wait until the end of the half-cycle to begin its response.

With the square waves assumed by Ringelman and Fenaroli (see reference 2 of Mr. Bourne's discussion), the solution of the difference equation becomes

$$N_g\varphi_n=\frac{E_s}{2}(1+E)\alpha-\frac{\Delta E_cR\pi}{2N}\left[1-\left(\frac{R-1}{R+1}\right)^n\right] \qquad (97$$

Either this equation or equation 91, or equation 91(A), predict an oscillatory response to a step of input for $R<1$, because of the alternating sign of the factor raised to the nth power. The oscillatory response can also be easily shown from Fig. 16. One can equate this to the integral of the discharging voltage between zero and the firing time t_n. The firing time is found without the requirement that the transient begin with zero output.

$$t_n=\frac{2}{\omega E_s(1+E')}N_g\varphi_n \qquad (98)$$

For the simplified case of zero initial output

$$t_n=t_0\left\{1-ER\left[1-\left(\frac{R-1}{R+1}\right)^n\right]\right\} \qquad (99)$$

which is equal to equation 16 of the reference. For the sine wave of supply voltage and rectified-sine-wave control

$$\cos\omega t_n=1-\frac{2}{E_s(1+E')}N_g\varphi_n \qquad (100)$$

which can be expanded in terms of equation 91(A) to give similar results.

In mode II, inspection of the graph corresponding to Fig. 17 gives the difference equation

$$N_g\varphi_{n+1}-\left(\frac{E'R-1}{E'R+1}\right)N_g\varphi_n=-\frac{2\Delta E_cR}{N(1+R)}+\frac{2E_s}{1+R}(1-\cos\alpha) \qquad (101)$$

This is the first-order difference equation with a nonconstant coefficient. The solution for a step of input voltage is

$$N_g\varphi_n=\frac{E_s(1+ER)}{1+R}(1-\cos\alpha)-\frac{\Delta E_c}{N}\left(\frac{R}{1+R}\right)(\cos\alpha'+\cos\alpha)\times\left[1-\left(\frac{E'R-1}{E'R+1}\right)^n\right] \qquad (102)$$

The author strongly recommends that the reader examine and corroborate the results of the various references. One additional reference is listed with two supplementary references on the origin, properties, and solution of difference equations.

REFERENCES

1. FORMULATION OF MAGNETIC AMPLIFIER CIRCUIT EQUATIONS, D. A. Schaefer. *AIEE Transactions*, vol. 74, pt. 1, Nov. 1955, pp. 619–27.

2. TRANSIENTS IN LINEAR SYSTEMS (book), M. F. Gardner, J. L. Barnes. John Wiley & Sons, Inc., New York, N. Y., chap. IX.

3. APPLIED MATHEMATICS FOR ENGINEERS AND PHYSICISTS (book), L. A. Pipes. McGraw-Hill Book Company, Inc., New York, N. Y., chap. X.

Direct-Water-Cooled Germanium Power Rectifier

R. E. WAHL
ASSOCIATE MEMBER AIEE

SINCE THE DEVELOPMENT of large-area germanium rectifier cells in 1952, over 20,000 kw of d-c power have been installed and are in service. An equally large amount of germanium rectifier units for converting alternating current to direct current are on order with the various manufacturers of this type of equipment. The first commercial germanium rectifier units used a germanium rectifier stack as shown in Fig. 1. The germanium rectifier cell was sealed with lacquer, making it partially impervious to moisture and atmosphere. This stack was then immersed in a tank containing perchlorethylene and connected with other stacks to form a number of 3-phase bridge rectifiers. The tray was then cooled by running tap water through a thin copper tube surrounding the germanium rectifier stacks. Later, the germanium rectifier stacks were installed in a tank and the perchlorethylene recirculated through several tanks and a water-to-perchlorethylene heat exchanger, to remove the losses from the germanium rectifier stacks. A recent tray using this construction is shown in Fig. 2 which forms part of a 65-volt d-c unit at 15,000 amperes.

During this time, work was going on to develop a hermetically sealed germanium rectifier cell. This construction, shown in Fig. 3, uses a metallized ceramic insulator of alumina. This development led directly to air-cooled hermetically sealed germanium rectifiers such as shown in Fig. 4, and to direct-water-cooled rectifier units, which is the subject of this paper.

Germanium Rectifier Cell

The heart of the rectifier is a small wafer of ultrapure single-crystal germanium which is sandwiched between an acceptor solder on one side and a donor solder on the other. Germanium is a semiconductor, which means that the material is neither a good conductor nor a good insulator. The atomic bonds of germanium atoms are readily broken by thermal agitation and some conduction occurs at room temperature, increasing greatly at temperatures over 100 C (degrees centigrade). Of much greater importance than these temperature effects is the asymmetrical conduction caused by impurities in the germanium crystal structure. Atoms with 5-valence electrons, such as phosphorus, arsenic and antimony, can enter the crystal-lattice in place of germanium atoms. Here four of the five electrons will enter the inner atomic bonds, the fifth electron being bound by only the slight electric charge on the nucleus. At room temperature, thermal agitation is sufficient to free these extra electrons so that they can wander through the lattice and conduct current. This is called *n*-type conduction (*n* for negative). Because these impurities provide conduction electrons, they are called donors.

If impurities such as boron, aluminum, gallium, and indium, are used, whose atoms have only 3-valence electrons, they produce so-called *p*-type conduction (*p* for positive) when they enter the crystal-lattice. Around each of these atomic impurities the structure will be locally incomplete, forming a trap for one more electron. An electron from a near-by bond can then jump into this trap, leaving a vacancy, or hole, behind. Another electron can migrate into the resulting vacancy, and so on in the process. This migration of electrons allows the hole to wander effectively about the lattice and in so doing to conduct electricity. This *p*-type conduction is in a direction opposite to that of electrons and is produced by impurities that are called acceptors.

In Fig. 3, then, the lower surface of the germanium alloyed with arsenic, a donor, produces *n*-type material, while the upper surface alloyed with indium produces *p*-type material in the same crystal. The junction between the *n* type and the *p* type is called a P-N junction. A P-N junction has a rectifying property, in that an electric current flows through it in one direction with ease but with great difficulty in the other direction. If the *p* region is made positive and the *n* region negative, holes and electrons are driven toward each other and recombined, each electron filling up one hole. The voltage required for this is low, being only enough to keep the current going. This is the forward direction or easy-flow direction of current. If the *p*-type region is made negative with respect to the *n*-type region, the electrons and holes move away from each other leaving the region close to the junction an insulator since it has no current carriers. This region is called the barrier layer. The inversely biased P-N junction, however, will conduct electricity in the inverse direction if carriers are introduced. Carriers are actually induced thermally by the breaking of bonds in the germanium which supplies conduction electrons and holes. This constitutes reverse current, also known as leakage current or saturation current. The remainder of the structure shown in Fig. 3 provides protection of the germanium rectifier junction against mechanical damage, moisture, and atmospheric conditions. It also provides a means for connecting electrical terminals to the device as well as a means for removing the heat generated by the forward current and the reverse current flowing through the junction.

Direct-Water-Cooled Germanium Rectifier Stack

The development of the hermetically sealed germanium rectifier cell made direct water-cooling of it possible. This complete rectifier stack is shown in Fig. 5. The construction enables the rectifier stack to be cooled by means of water flowing between the seven projections on one side of the stack. Fig. 6 shows these projections with a Buna N rubber gasket to seal one surface of the stack when mounted, as shown in Fig. 7. In a close-up view, Fig. 8, shows the rectifier stack mounted to a short section of stack support. Water flowing through this stack support flows past the projecting pins and removes heat generated by the rectifying junction in a very efficient manner. This is shown in Fig. 9, which is a plot of cell rise versus watts dissipated in the junction. It shows approximately 1/4 C rise per watt dissipated in the junction. Other structures may exhibit as high as 1 C rise

Vol. 7
Pt. 1
1956

Paper 56-976, recommended by the AIEE Metallic Rectifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Chicago, Ill., October 1–5, 1956. Manuscript submitted July 6, 1956; made available for printing August 16, 1956.

R. E. Wahl is with the General Electric Company, Lynn, Mass.

Fig. 1. Type-RA3 industrial germanium rectifier cell

per watt or higher. This curve was run experimentally by causing a direct current to flow through the rectifying junction in the forward direction and measuring d-c volts and d-c amperes across the stack. The cell temperature rise was measured by having a thermocouple imbedded in a small hole located a few thousandths of an inch away from the rectifying junction.

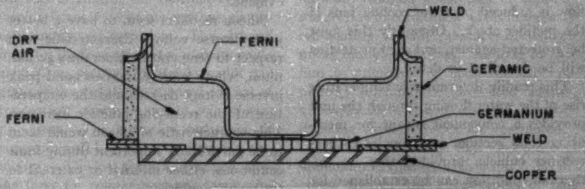
Further tests show that this method measures the junction temperature to within about 2 C below the actual junction temperature. This was determined by a plot on semilog paper of saturation current versus junction temperature with the rectifier stacks in a temperature-controlled oven. Here, the temperature of the junction is known, since the thermal gradient between the junction and the measuring point is zero, after the stack has reached constant temperature of oven and since no internal losses are present in the stack. This curve, then, was used with the stack mounted on its stack support and a plot was made of watts dissipation in the junction versus thermocouple temperature and temperature by saturation current.

To ensure that sufficient time was allowed between points on the foregoing curves, the thermal time constant of the rectifier stack was determined experimentally, as shown in Fig. 10. This indicates a thermal time constant of about 30 seconds, the point at which the junction has reached 63.2% (per cent) of its final steady-state temperature rise. Thus, in five time constants or about 2½ minutes, the junction will have reached its steady-state temperature rise to better than 99%.

The static d-c characteristic of a cell is shown in Fig. 11, which shows a forward voltage drop of approximately 0.6 volt at 83⅓ amperes direct current. This corresponds to approximately a 50-watt forward loss, while the reverse losses constitute an additional 20%, making a total of approximately 60 watts in the junction to be dissipated. It is the removal of this 60 watts from the germanium wafer that constitutes the major problem in spite of its being a small amount of power. This is one of the main reasons that the principle of direct water cooling was developed, eliminating, as it does, as many interchanges of heat between the rectifying junction and the coolant which finally removes the heat generated. Limiting the junction temperature to below 100 C, which is a maximum permissible temperature, requires that the junction then be cooled with a reasonable engineering safety factor to about 65 C maximum. This means that if the stacks are to be cooled by a maximum water temperature of 35 C, which is permitted to rise 5 C as it goes through the series of rectifier stacks connected to the stack support, the remaining difference of 25 C is all that can be allowed for the rise of the junction above water at any one point.

Fig. 2. Model no. 6RA13DW1 germanium rectifier tray rated 1,667 amperes, pumped perchlorethylene-cooled, containing 20 6RA3-type rectifying devices and mounting current-limiting fuses and triggers. Outside view

The efficient removal of heat from the rectifier stack is, among other things, a function of water velocity past the projecting pins, shown in Fig. 12. This curve shows that at about 3 gpm (gallons per minute) with three extrusions connected in parallel, cells are cooled to within approximately 2 C of what they will be for an infinite water velocity through the cooling surfaces. Fig. 13 shows the temperature rise of water



HERMETICALLY SEALED GERMANIUM RECTIFIER CELL

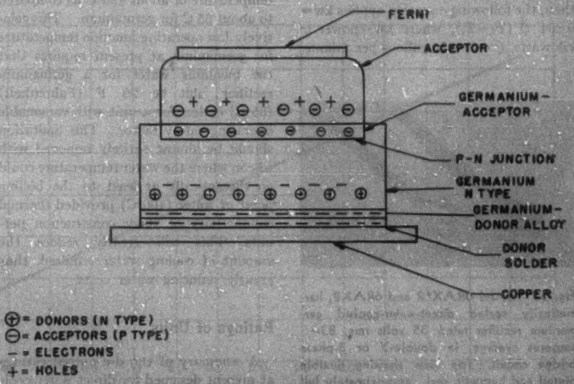


Fig. 3. Enlarged cross section of germanium rectifier cell

Fig. 4. Model 6RA14 (A-D) B4 hermetically sealed forced-air-cooled germanium rectifier, rated up to 65 volts rms, 83-$^1/_3$ amperes average, in double-Y or 3-phase bridge circuit; 7-fin design for reduced velocity cooling

flowing through three of the cooling extrusions in parallel versus gallons per minute of water. This curve shows that at about 3 gpm of water the temperature rise in the water is about 3 C. Since temperature rise of the water is inversely proportional to the rate of flow of water, the water rise would be 6 C at 1$^1/_2$ gpm. At one half of this flow, or 3/4 gpm, the temperature rise of the water would be 12 C. From this and the appearance of the curve it is apparent that a runaway condition exists for both the cooling water and the rectifier junction, as water flow is reduced past the cooling pins of the rectifier stack. Obviously this must be protected against, and such protection will be described in detail.

This readily determinable temperature rise of the water flowing through the unit provides a convenient means for measuring total rectifier junction losses in a rectifier cubicle, provided that an isothermal condition can be established for the stack supports. This can be done by having the average water temperature going through the unit equal to ambient in which the unit is immersed. Here, then, the following equation applies kw = 0.264 Q $(T_2 - T_1)$, where kw = power in kilowatts, Q = flow in gallons per minute, T_2 and T_1 = outlet and inlet temperatures in degrees centigrade. The temperature difference may be determined accurately by using relatively cool water and permitting a relatively large temperature rise in the water, since quantity of water may be measured accurately by flow into a standard volume container over a given length of time or by use of a precision flow-indicating device.

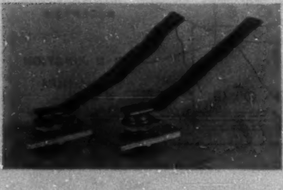


Fig. 5. Model 6RAX12 and 6RAX8, hermetically sealed direct-water-cooled germanium rectifier rated 65 volts rms, 83-$^1/_3$ amperes average, in double-Y or 3-phase bridge circuit. Top view showing flexible pigtail and hermetic seal; approximately full size

Silicon rectifier stacks can be applied directly to this type of direct water cooling by a simple substitution of the silicon rectifier for the germanium rectifier. At present, the highest current rating for a silicon rectifier is about 50 amperes average. Fortunately, silicon cells can be produced with considerably higher peak inverse voltage capacity so that while the forward current loss per cell is approximately twice that for germanium only one-half or fewer silicon rectifier cells will be required for higher voltage units than in the present germanium rectifier units. This means that for voltages above about 80 volts direct current, where two or more germanium rectifiers are required in series, the value of the silicon rectifier will be self-evident.

Silicon rectifiers seem to have a better peak inverse voltage characteristic with respect to temperature than does germanium, whose leakage increases and peak inverse voltage decreases as the temperature of the rectifying junction increases. This characteristic of silicon would seem to be particularly important during fault conditions, either internal or external to the rectifier cubicle.

The major advantage of silicon over germanium appears to be a considerably higher permissible operating junction temperature of about 150 C as compared to about 65 C for germanium. This relatively low operating junction temperature for germanium at present requires that the incoming water for a germanium rectifier unit be 95 F (Fahrenheit) (35 C) or less for a unit with reasonable operating safety factor. This limitation should be almost entirely removed with silicon where the water temperature could go theoretically at least to the boiling point of water (100 C) provided thermal conductivity of stack construction permits. This will greatly reduce the amount of cooling water required, thus greatly reducing water costs.

Fig. 6. Model 6RA20DW1, hermetically sealed direct-water-cooled germanium rectifier rated 65 volts rms, 83-$^1/_3$ amperes average in double-Y or 3-phase bridge circuit. View of underside of rectifier showing cooling pins and O-ring gasket

Ratings of Units

A summary of the d-c output ratings at present designed for direct water cooling is as follows:

Steel-mill rated units
- 100% rated output continuous
- 125% rated output ADC (amperes direct current) (2 hours)
- 150% rated output ADC (1 minute)
- 12, 15, 20, 25, 30 VDC (volts direct current) at 4,000, 5,000, 7,500 ADC

Electrochemical or metallurgical with transformer in rectifier cubicle
- 100% rated output continuous
- 150% rated output ADC (one minute)
- 50, 65, 75 VDC at 2,000, 2,500, 3,000, 4,000, 5,000 and 6,000 ADC
- 100, 130, 150 VDC at 1,000, 1,500, 2,000, 3,000 ADC

Electrochemical or metallurgical power supplies with transformer external to rectifier cubicle
- 100% rated output continuous
- 150% rated output ADC (one minute)
- 50, 65, 75 VDC at 8, 10, 12, 14, and 16 kilo-ADC
- 100, 130, 150 VDC at 5, 6, 7, 8, 10, 12, 14, 16 kilo-ADC

This listing shows units with a kilowatt output rating from 48 to 225 covering a voltage range from 12 to 30 volts direct current. These units use circuit number 49 in the ASA standard.[1] The wye primary on the transformer has a shifting neutral which provides 120-degree cell conduction in the rectifying elements. This circuit permits elimination of the interphase transformer used in circuit 45 in the same standard, thereby eliminating its losses and regulation. However, a corresponding change in transformer design must be dealt with.

The electrochemical units use either circuits 23, 24, 25, or 26 in the foregoing standards with kilowatt ratings from 100 to 450, where the transformer is included in the rectifier cubicle. This type of unit is preferred where a double voltage transformation is desirable as, for example, where an existing substation providing one of the standard voltages up to about 600 volts is already available. Ratings of units from 400 to 2,400 kw output power can be obtained in a single cubicle which uses a transformer external to the rectifier cubicle. This type of unit is desirable when it is more economical to use a single

Fig. 7 (left). Germanium rectifier cell assemblies mounted on silver-plated aluminum extruded H-section tube. Tube will conduct both current and cooling water

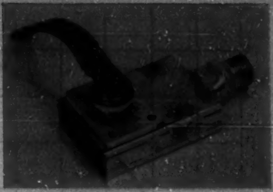


Fig. 8. Model 6RAX8, hermetically sealed direct-water-cooled germanium rectifier rated 65 volts rms, 83-1/3 amperes average, in double-Y or 3-phase bridge circuit. View showing rectifier in position; hollow, water-cooled bus bar

magnetic voltage conversion from a-c power systems of 2,400 up to about 15,000 volts. This type of installation is more economical in large power installations and where a new installation is being made.

The higher voltage electrochemical units, i.e., 100-, 130-, or 150-volt direct current, use two germanium rectifier stacks in series to provide sufficient inverse voltage rating. While the lower voltage units, i.e., 50-, 65-, and 75-volt d-c, can be put into series to provide 100-, 130-, and 150-volt direct voltages and higher, it is more economical to series stacks mainly because of savings in bus conductors carrying a-c power to the rectifying stacks. Some savings also accrue because the d-c outputs in the series stack units are essentially one output and need not be tied in series.

While the chart does not show units rated higher than 150 volts direct current, these units may be connected in series to get up to 600 volts direct current until such time as higher direct-voltage units are designed. Further experience with the presently designed units is required before voltage may be extended beyond this.

Electrochemical Problems

Because aluminum is basically less expensive as a current-carrying part than copper, considerable effort was devoted to making the stack support and germanium cell holder out of this element. It was found that the stack support could be an aluminum extrusion and the cell holder a die casting. It also became evident that the prevention of corrosion of these aluminum parts was extremely difficult because of their contact with water.

The main problem in using aluminum parts is to control the thermal-dynamic driving forces so as to prevent any significant amount of corrosion. This driving force for the formation of aluminum oxide at 25 C approaches 7,000 calories for one gram of the metal as compared to that of copper into cuprous oxide of 300 calories per gram. The corrosion resistance of copper is well known, as is that for aluminum. The low thermal-dynamic driving force for copper accounts for its inherent inactivity, but certainly does not account for aluminum's apparent inactivity.

Oxide Film

The invisible oxide film which forms on aluminum at room temperature has made possible the use of aluminum for high-voltage transmission-line cables since the turn of the century; its use also includes aluminum storm windows, entire building facades, entire aircraft fuselage and wing structures, as well as the commonplace aluminum pots and pans. This naturally formed oxide film is approximately one micron thick. It is inherently self-healing in the presence of oxygen under many conditions. As long as this film is unbroken or spontaneously repaired, aluminum is protected against atmospheric corrosion, and the thermodynamic driving forces are restrained.

This invisible oxide film also protects aluminum in contact with water. To increase further the effectiveness of this oxide film, the aluminum may be anodized by an electrolytic process which builds up this film up to about one mil in thick-

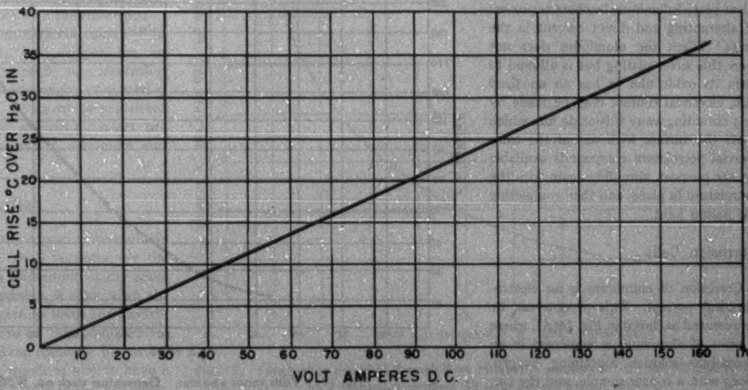


Fig. 9. Volt-amperes versus cell rise degrees centigrade above H_2O ambient 4.5 gpm H_2O, series tubes germanium stack no. W56

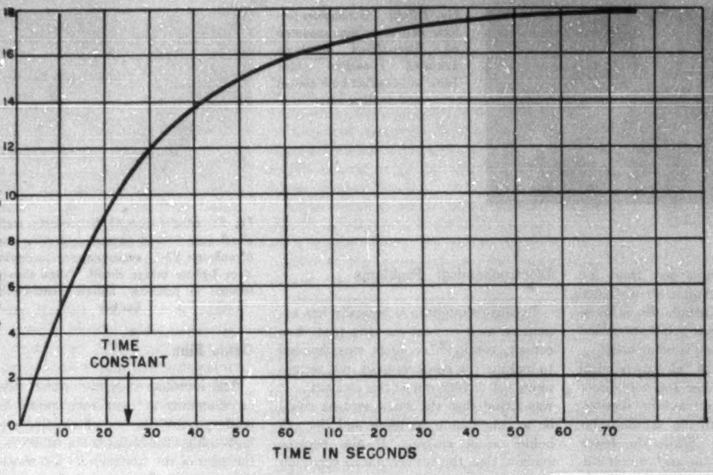


Fig. 10. H_2O-cooled germanium 6RAX8. Thermal time constant. Thermocouple in drilled hole in cell

ness. This aluminum oxide, however, is a semiconductor and has a relatively high resistance to the flow of current. Because of this it is desirable to make electric connections to the aluminum where the anodic film has been removed or prevented from forming, or to use another type of surface protection which provides good electrical conductivity.

In the past few years, the quality and cost of silver-plating aluminum surfaces has been greatly improved and it can now be used to restrain the thermodynamic driving forces and also to provide an electrical contact surface to the aluminum. Several large installations of metallic rectifiers have been in operation for years using silver-plated aluminum bus bars as carriers of alternating and direct current in the units. When the aluminum does not have this silver plating but is allowed to form its oxide film or has an anodized film, electrical contact may be made by first abraiding away this oxide film which coats the surface with one of the commercial petrolatum compounds available for the purpose, abraiding again with the petrolatum in place, and then completing the bolted joint.

Corrosion Cells

Corrosion of aluminum is an electrochemical process. This process can be represented as shown in Fig. 14(A), where a piece of aluminum is immersed in an electrolyte solution containing oxygen, along with a copper sample, and the two are connected together electrically. In this case, the aluminum corrodes proportional to the number of electrons which flow from the aluminum to the copper in the external circuit. The electrode reactions are shown for both the anode and cathode in Fig. 14.

In the foregoing example it would be easy to stop corrosion of the aluminum by merely disconnecting the two from one another electrically. Unfortunately, this is not the usual case, and anodes and cathodes may occur as shown in Fig. 14-(B). Here, copper has been deposited onto aluminum and corrosion results, as shown, with the same anode and cathode reactions taking place. This action causes the metal to corrode in spots, and it is this localized corrosion which causes the most trouble, since it may lead to perforation. Fortunately, the oxide film may build up in the corroded areas and prevent perforation, and the inevitable leakage of electrolyte would follow.

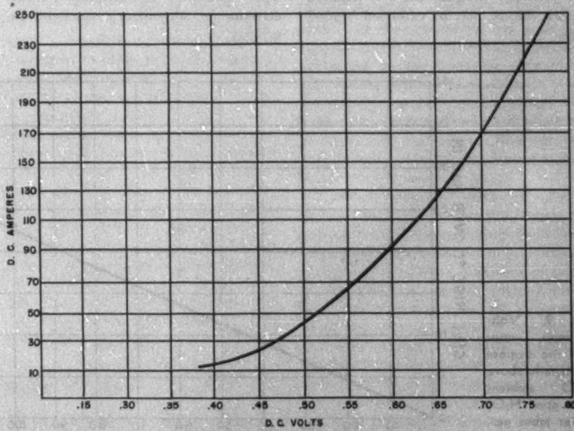


Fig. 11. Volts versus amperes. Germanium stack no. W56

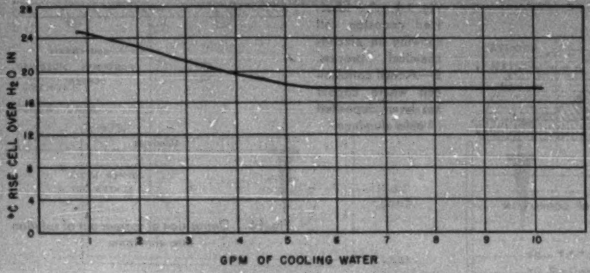


Fig. 12. Cell temperature rise versus gpm coolant

This localized attack of aluminum can be expected when the following four requirements are satisfied:

1. A cathode reaction can take place in conjunction with the anodic oxidation of aluminum, which can give rise to an electromotive force.
2. An electrolyte joins anode and cathode.
3. Anode and cathode are in metallic contact.
4. The natural oxide film on aluminum is broken or destroyed and a protective film cannot reform.

Fig. 15 shows the results which follow when aluminum and copper are joined so as to meet these four requirements. Here corrosion could have been prevented by keeping the joint dry or preventing moisture from gaining entrance to the junction by a suitable coating. Where moisture is unavoidable, corrosion may be prevented by measures which prevent copper from being transferred to the aluminum. This occurs as shown in Fig. 16. The best way to prevent this, of course, is not to have copper in the electrolite or in contact with the electrolite.

Fig. 17 shows another corrosion cell in which aluminum and copper are in metallic contact with an electrolite between them which results in corrosion of the aluminum. This can be prevented by removing the electrical contact between the two elements, aluminum and copper.

Cathodic Protection

In Fig. 16(A) the copper is the cathode and is protected from corrosion by the corroding of the aluminum in the same electrolyte. Corrosion is always at the anode, where electrons leave the electrolite. Metallic ions at this point are released into solution resulting in rusting or pitting. The anode is said to have a higher solution potential than the cathode, i.e., a greater tendency to go into solution. This can be illustrated quickly by the experiment shown in Fig. 18, where a silver coin is connected through a milliammeter to a nail. Here the nail will act as an anode, the current will flow, and a rust cloud will appear surrounding the nail. The silver coin is cathodic and is said to be cathodically protected. In this case, the nail gradually erodes away and is said to be sacrificial. The milliammeter indicates a galvanic current flowing.

If this galvanic current can be neutralized, overcome, or stopped by an equal or greater flow of current in the opposite direction, then the anodic areas of the metal to be protected become cathodic to the auxiliary anodes. This is shown in Fig. 19, where two identical nails are suspended in an electrolyte and connected to an external battery. The nail connected to the positive terminal of the battery is the anode, and a rust cloud soon appears surrounding it in the solution. The other nail connected to the negative terminal of the battery is cathodically protected.

Frequently it is not desirable to have an anode which is sacrificed since it may contaminate the electrolyte. This contamination can be prevented by using in place of a sacrificial anode a permanent, nonsacrificial anode made of a platinum wire or a stainless steel rod clad with platinum foil. This functions as follows, when oxygen ions are attracted to the positive or anodic terminal where they are released but can combine chemically with an iron anode to form oxide or rust. If a platinum wire or terminal is used for the anode, the rust cloud disappears because there is no longer any chemical union of oxygen with the metal and the anode, thus the platinum remains permanently and is nonsacrificial. Hydrogen is released at the negative or cathodic terminal, and does not join chemically with the cathode but rises as bubbles to the surface. Fig. 20(B) shows how a permanent nonsacrificial anode made of platinum may be used to protect an aluminum shell when a difference in potential is applied between the platinum anode and the aluminum shell.

In this project, then, the main methods used to prevent corrosion of the aluminum extrusion were the use of cathodic protection for over-all protection of the parts, and the use of surface films as described to prevent local action, as shown in Figs. 21 and 22. The combination of

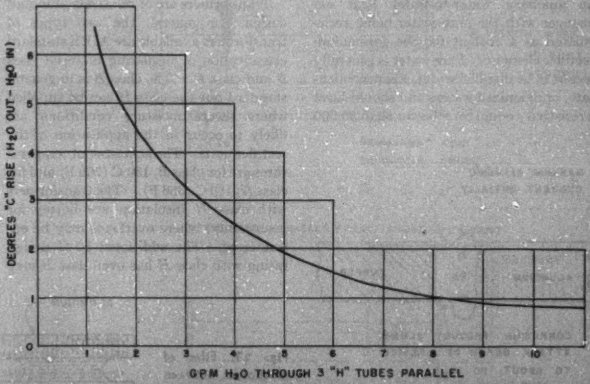


Fig. 13. Temperature rise H_2O through three H-tubes in parallel; 6-1-6 Y PRI YY SEC 9/16-inch germanium rectifier; 58 watts per cell average losses

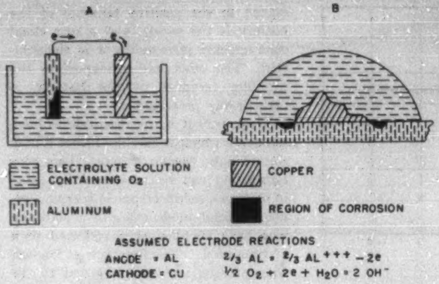

A
B
ELECTROLYTE SOLUTION CONTAINING O_2
COPPER
ALUMINUM
REGION OF CORROSION
ASSUMED ELECTRODE REACTIONS
ANODE = AL $2/3\ AL = 2/3\ AL^{+++} - 2e$
CATHODE = CU $1/2\ O_2 + 2e + H_2O = 2\ OH^-$

Fig. 14. A—Idealized corrosion cell showing an electrochemical process. B—Actual corrosion cell where copper has been deposited onto aluminum

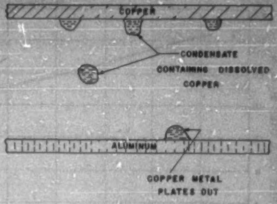


Fig. 16. Deposition of copper out of solution into aluminum

these two methods of protection where reasonable care is taken to prevent breakage of the protective surface film provides excellent corrosion resistance to waters having pH values between about 4.5 and 8.5. The presence of chlorides, particularly in conjunction with heavy metal ions, is generally deleterious and is also prevented by separation of these two components electrically. This separation of heavy metal ions from the aluminum could be provided to as great an extent as possible by having fluid control elements separated from the aluminum extrusion by lengths of polyvinyl chloride plastic pipe.

Tap water can usually be used directly with this type of unit provided the pH is within the limits as previously specified and its minimum resistivity is 2,000 ohm-centimeters. Tap water here means drinking water. Pure water, of course, can be used where satisfactory tap water is not available. This requires the use of an auxiliary water-to-water heat exchanger with the pure water being recirculated as a coolant for the germanium rectifier elements. Pure water is generally meant to be distilled water, steam condensate, or deionized waters and should have a resistivity equal to or better than 30,000 ohm-centimeters. Whenever there is a question as to the suitability of available cooling water, the recirculating system should be used with periodic resistivity measurements being made on the pure water side.

Because water supplies may vary considerably from locality to locality and from time to time it was decided to use copper stack supports and cell holders. Direct water-cooling is permissible where chloride content, heavy metal ions, pH and solids are below American Standards Association limits. If these limits are exceeded a closed recirculating system is required. Protection for these copper parts is required against anodic corrosion with the use of either nonsacrificial platinum anodes (or targets, as they are called) or stainless steel anodes which very slowly erode away.

Transformers

Transformers are of relatively standard design. In general, the two types of transformers available are AIEE standard classification of insulating material class *B* and class *H*. The class *B* is in general standard but the class *H* can be supplied where severe moisture conditions are likely to occur in the application of the rectifier units. The hottest-spot temperatures are for class *B*, 130 C (266 F) and for class *H*, 180 C (356 F). The transformers with class *H* insulation are better for heavy duty, where overloads may be encountered. The additional 50 C design rating with class *H* has over class *B* may be used to provide longer transformer service or reduction of weight of unit. The cores used in the 3-phase power transformers are grain-oriented silicon steel assembled into a core-type transformer. This grain-oriented steel has high saturation flux density, and lower core losses and exciting volt-amperes than were previously available from regular silicon steel. These characteristics enable operation at higher inductions with subsequent reductions in core cross section and hence weight of transformer. Each silicon steel strip is treated on both sides with a modified high-temperature coating which provides exceptionally good interlamination resistance with a negligible effect on space factor.

Transformer windings are form-wound on arbors of the proper size, insulated, and braced to withstand the mechanical stress between the current-carrying conductors because of overloads or short circuits on the rectifier units. The primary windings are of conventional construction with the use of copper or aluminum as the current-carrying part. Secondaries are of an improved design based on strip copper or strip aluminum wound spirally. This design is especially suitable where heavy d-c winding currents are required on a transformer kva basis as compared to the usual design of power transformers. The aluminum or copper strip is almost a complete length of the transformer core window with provision for proper yoke insulation requirements being met. High space factors are maintained on the d-c windings by using a Mylar polyester film

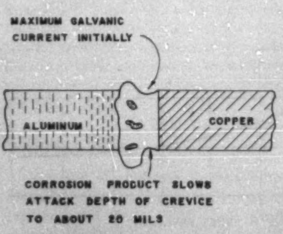

MAXIMUM GALVANIC CURRENT INITIALLY
ALUMINUM
COPPER
CORROSION PRODUCT SLOWS ATTACK DEPTH OF CREVICE TO ABOUT 20 MILS

Fig. 15. Butt-welded aluminum and copper joint 30 days in salt spray

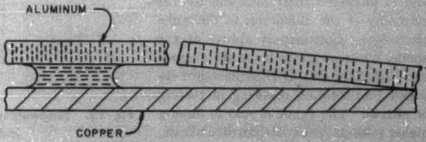


Fig. 17. Film of electrolyte between aluminum and copper in metallic contact

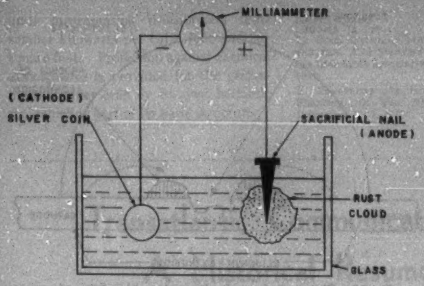


Fig. 18. Current flow from rusting nail

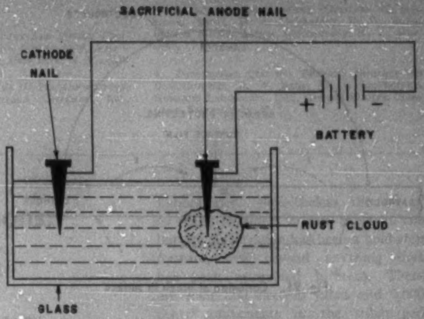


Fig. 19. Battery makes nails unlike in solution potential

bonded to highly purified asbestos, which provides the great mechanical strength required as well as high electrical breakdown characteristics.

Transformers are normally checked for ratio, polarity, and phase relations, no-load loss, exciting current, load loss, impedance, dielectric tests (1. applied potential, and 2. induced potential), and temperature rise on one unit of a new design. Other special tests which may be run where required by customers are temperature rise on duplicate units, impulse test, noise level test, zero sequence impedance test, and insulation power factor test. Each transformer also has an overtemperature thermoswitch imbedded in its center leg near the estimated hot-spot point in the winding to provide protection against short circuits and excessively long overloads which might cause winding insulation to deteriorate more rapidly than is desired.

Class *B* transformers are insulated with a general-purpose clear baking varnish with a phenolic drying oil base. The complete transformer is vacuum-impregnated to provide proper surface coating, and filling of spaces in porous insulation and interstices between copper turns; to build up film insulation upon conductors, coils, and coil assemblies; and to bond coils and coil assemblies together for additional electrical and mechanical strength. Class *H* transformers use silicon resin as the insulating varnish. The use of silicon varnish improves most of desired characteristics mentioned in the foregoing.

Protection of Units

Rectifier units are normally protected against d-c short circuit, rectifier cell failure, coolant overtemperature, loss of coolant, minimum flow, voltage surges detrimental to germanium cells and d-c overload. D-c short-circuit protection is provided but, in general, germanium rectifier applications are limited to loads where d-c short circuits are extremely unlikely or impossible. This protection is provided by a-c circuit breakers coordinated with d-c overload protective devices. Instantaneous breaker trip elements are set down to approximately four times the a-c rating of the rectifier unit. Air circuit breakers are, in general, used for units of this type, their particular selection being determined by circuit voltage and interrupting rating required, and load current-carrying capacity.

A small percentage of germanium rectifier cells fail in such a manner that they lose their ability to prevent the flow of current in the reverse direction and become essentially short circuits. Protection against this is provided by high-speed fuses that can withstand less I^2T than the germanium rectifier cell, thus providing protection for good cells against possible short circuit of any potentially defective cell. This type of protection is similar to arc-back protection required in mercury arc rectifiers.

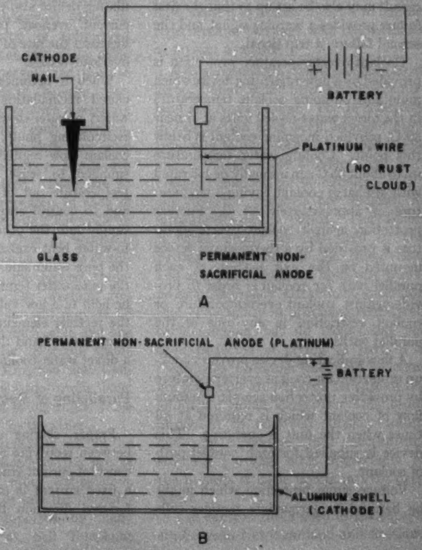


Fig. 20. A—Nonsacrificial anode cathodically protects nail. B—Permanent nonsacrificial anode (platinum)

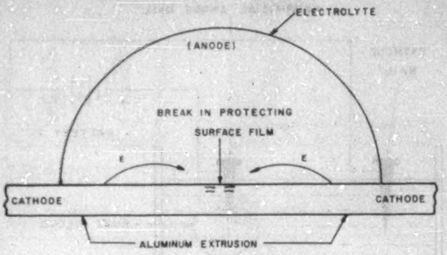


Fig. 21. Corrosion at break in surface

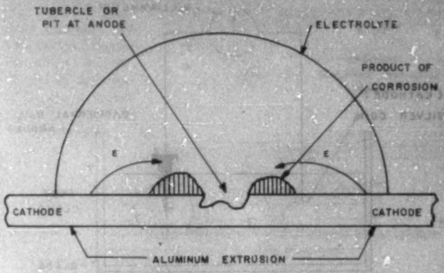


Fig. 22. Pitting attach, result of corrosion at break in surface

The use of these germanium power fuses necessitates some type of monitoring system for these fuses. If such a system were not provided, loss of some of the rectifying elements could cause loss of additional rectifying elements. In the early stages of development of this type of germanium rectifier unit a solenoid-operated stepping relay was used to check each fuse at regular intervals and to provide either a trip or warning signal when a burned fused was discovered. Another type of fuse-monitoring circuit consists of fuses connected across the main fuses. When the main fuse burns, the trigger fuse also burns and releases a plunger which closes a signal circuit which may be used either to warn or to trip. This circuit may also be set up so that the first failure provides a warning signal, and the second failure, a trip signal.

Coolant overtemperature protection is provided by a flow-regulating device which maintains outgoing coolant temperature in the direct water-cooled units at or near 30 C, provided incoming coolant is below this point. Cooling water temperature up to about 35 C can normally be tolerated with discharge coolant having a temperature of approximately 40 C. Further protection against coolant overtemperature is provided by a thermostatic device mounted on the exit portion of each parallel path for coolant in order to provide against coolant overtemperature or coolant restriction in any one of the parallel paths.

A flow-indicating device provides back-up protection for coolant overtemperature by providing protection against minimum flow of coolant which is required at all times when the unit is operating. This device is supplied for each parallel path of coolant.

Voltage surge protection which might be harmful to germanium rectifier cells is provided by Thyrite resistors and surge suppression capacitors connected as mentioned previously. These protect against a-c power supply surges, d-c circuit induced surges, and internally produced surges.

D-c overload protection is provided by standard commercial induction disk-type relays or plunger-type relays. The particular type of relay selected is determined by the nature of the d-c load and whether overload characteristics are required. For example, a d-c load may require 125% d-c load current for one minute or even 150% d-c load current for one minute. This requires proper co-ordination between all of the protective devices involved in the system. The flow-regulating device also provides a measure of current overload protection in that it increases the rate of coolant flow when an overload occurs.

When it becomes necessary to use a closed recirculating system with pure water, a water-to-water heat exchanger, recirculating pump, and reservoir with coolant flow regulator will normally be supplied to sense pure water coolant temperature out of the heat exchanger and to control entry water from water supply into the other side of the water-to-water heat exchanger. Since flow on the pure water side is at a constant rate, the pure-water temperature gradient can be held to a low value, thus maintaining the rectifier elements within a narrower temperature band than can be done with a direct water-cooling system.

Paralleling of Rectifier Cells

Proper sharing of the load current between rectifying cells is provided by a constant-current grading method in which a fixed current is caused to flow in the forward direction through the rectifier element, and the voltage drop across the cell and its supporting structure is measured. Installation of rectifier cells in groups having a voltage drop within a very narrow band ensures that each rectifying cell will be carrying its proper amount of load current. Cells are also graded in accordance with their peak inverse voltage characteristics to ensure their being operated properly on the particular circuit and the particular voltage rating in which they are being used.

Regulation of D-C Output

Two types of direct voltage control can normally be provided since germanium rectifier units cannot be regulated by phase control as mercury arc rectifiers can be. Induction voltage regulators or saturable reactors are the two common types. Saturable reactors provide a faster response in general than do induction voltage regulators or the rather infrequently used tap-changing transformer control, the latter being of either the no-load or on-load type.

Improvements

Major improvements accomplished by this design are ease of cell replacements, reduction of cost per kilowatt output and inertness to chemical atmospheres. Rectifier cells can be replaced readily by unbolting four bolts holding the rectifier stack in place on its cooling tube. The cost per kilowatt has been reduced drastically by the new design which reduces floor space required on most of the units as well as providing improved cooling for rectifier junctions, thus providing longer life. Some of the transformerless units have almost 10 kw output per cubic foot of volume occupied and up to 60 kw per square foot of floor space required. The units incorporating transformers show smaller kilowatts per cubic foot and per square foot. Protection against chemical atmospheres is provided for the germanium rectifier cells by its own hermetic seal which is part of its basic structure.

References

1. Pool Cathode Mercury-Arc Power Converters. *ASA Standard No. C34.1—1949*, American Standards Association, New York, N. Y., Jan. 1949.

2. Electrons and Holes in Semiconductors, Shockley. D. Van Nostrand Company, Inc., Princeton, N. J., 1950.

3. Aluminum for Copper; Greater Corrosions, Risks, H. A. Liebhafsky, E. W. Balis. *General Electric Review*, Schenectady, N. Y., Mar. 1953.

4. Protecting Cooling Systems Subject to Electrolysis, E. W. Balis, H. A. Liebhafsky. Research Laboratory, 1949, General Electric Company, Schenectady, N. Y.

Transatlantic Communications: A Historical Résumé

MERVIN J. KELLY SIR GORDON RADLEY

THE PAPERS that follow describe the design, manufacture, and installation of the first transatlantic telephone cable system in all its component parts, including the connecting microwave radio-relay system in Nova Scotia. The purpose of this introduction is to set the scene in which this project was undertaken and to discuss the technical contribution it has made to the development of world communications.

Electrical communication between the two sides of the North Atlantic started in 1866. In that year, the laying of a telegraph cable between the British Isles and Newfoundland was successfully completed. Three previous attempts to establish transatlantic telegraph communication by submarine cable had failed. These failures are today seen to be the result of insufficient appreciation of the relation between the mechanical design of the cable and the stresses to which it is subjected as it is laid in the deep waters of the Atlantic. The making and laying of deep-sea cables was a new art and designers had few experiments to guide them.

During the succeeding 90 years, submarine telegraph communication cables have been laid all over the world. Cable design has evolved from the simple structure of the first transatlantic telegraph cable—a stranded copper conductor, insulated with gutta-percha and finished off with servings of jute yarn and soft armoring wires—to the relatively complex structure of the modern coaxial cable, strengthened by high-tensile steel armoring for deep-sea operation. The coaxial structure of the conducting path is necessary for the transmission of the wide frequency bandwidth required for many telephone channels of communication. The optimum mechanical design of the structure for this first transoceanic telephone cable has been determined by many experiments in the laboratory and at sea. As a result, the cable engineer is confident that the risk of damage is exceedingly small even when the cable has to be laid and recovered under conditions which impose tensile loads approaching the breaking strength of the structure.

The great difference between the transatlantic telephone cable and all earlier transoceanic telegraph cables is, however, the inclusion of submerged repeaters as an integral part of the cable at equally spaced intervals and the use of two separate cables in the long intercontinental section to provide a separate transmission path for each direction. The repeaters make possible a very large increase in the frequency bandwidth that can be transmitted. There are 51 of these submerged repeaters in each of the two cables connecting Clarenville in Newfoundland with Oban in Scotland. Each repeater provides 65 decibels of amplification at 164 kc, the highest transmitted frequency. The working frequency range of 144 kc will provide 35 telephone channels in each cable and one channel to be used for telegraph traffic between the United Kingdom and Canada. Each cable is a 1-way traffic lane, all the "go" channels being in one cable and all the "return" in the other.

The design of the repeaters used in the North Atlantic is based on the use of electron tubes and other components, initially constructed or selected for reliability in service, supported by many years of research at Bell Telephone Laboratories. Nevertheless, the use of so many repeaters in one cable at the bottom of the ocean has been a bold step forward, well beyond anything that has been attempted hitherto. There are some 300 electron tubes and 6,000 other components in the submerged repeaters of the system. Many of the repeaters are at depths exceeding 2,000 fathoms ($2^1/_4$ miles) and recovery of the cable and replacement of a faulty repeater might well be a protracted and expensive operation. This has provided the incentive for a design that provides a new order of reliability and long life.

On the North Atlantic section of the route, the repeater elements are housed in flexible containers that can pass around the normal cable-laying gear without requiring the ship to be stopped each time a repeater is laid. The advantages of this flexible housing have been apparent during the laying operations of 1955 and 1956. They have made it possible to continue laying cable and repeaters under weather conditions which would have made it extremely difficult to handle rigid repeater housings with the methods at present available.

A single connecting cable has been used across Cabot Strait between Newfoundland and Nova Scotia. The 16 repeaters in this section have been arranged electronically to give both-way amplification and the single cable provides "go" and "return" channels for 60 circuits. "Go" and "return" channels are disposed in separate frequency bands. The design is based closely on that used by the British Post Office in the North Sea. Use of a single cable for both-way transmission has many attractions, including that of flexibility in providing repeatered cable systems, but no means has yet been perfected of laying as part of a continuous operation the rigid repeater housings that are required because of the additional circuit elements. This is unimportant in relatively shallow water, but any operation that necessitates stopping the ship adds appreciably to the hazards of cable laying in very deep water.

The electron tubes used in the repeaters between Newfoundland and Scotland

---

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957.

Mervin J. Kelly is with the Bell Telephone Laboratories, New York, N. Y., and Sir Gordon Radley is with the British Post Office, London, England.

are relatively inefficient, judged by present-day standards. They have a mutual conductance of 1,000 micromhos. Proved reliability, lower mechanical failure probability, and long life were the criteria that determined their choice. Electron tubes of much higher performance with a mutual conductance of 6,000 micromhos are used in the Newfoundland-Nova Scotia cable, and it is to be expected that long repeatered cable systems of the future will use electron tubes of similar performance. This will increase the amplification and enable a wider frequency band to be transmitted, thus assisting provision of a greater number of circuits. If every advantage is to be taken of the higher performance tubes, it will be necessary to duplicate (or parallel) the amplifier elements of each repeater, in the manner described in a later paper, in order to assure adequately long trouble-free performance. This has the disadvantage of requiring the use of a larger repeater housing.

During the 3 years that have elapsed since the announcement in December 1953, by the American Telephone and Telegraph Company, the British Post Office, and the Canadian Overseas Telecommunication Corporation, of their intention to construct the first transatlantic telephone cable system, considerable progress has been made in the development and use of transistors. The low power drain and operating voltage required will make practicable a cable with many more submerged repeaters than at present. This will make possible a further widening of the transmission band which could provide for more telephone circuits with accompanying decrease in cost per speech channel or the widened band could be utilized for television transmission. Much work, however, is yet to be done to mature the transistor art to the level of that of the thermionic electron tube and thus insure the constancy of characteristics and long trouble-free life that this transatlantic service demands.

The present transatlantic telephone cable whose technical properties are presented in the accompanying papers, however, gives promise of large reduction in costs of transoceanic communications on routes where the traffic justifies the provision of large traffic capacity repeatered cables. The thirty-six 4-kc channels which each cable of the 2-way system provides, are the equivalent of at least 864 telegraph channels. A modern telegraph cable of the same length without repeaters would provide only one channel of the same speed. The first transatlantic telegraph cable operated at a much slower speed, and transmitted only three words per minute. The greater capacity of future cables will reduce still further the cost of each communication circuit provided in them. Such considerations point to the economic advantage, where traffic potentials justify it, of providing broad-band repeatered cables for all telephone, telegraph, and teletypewriter service across ocean barriers.

The new transatlantic telephone cable supplements the service now provided by radio telephone between the European and North American continents. It adds greatly to the present traffic-handling capacity of this service. The first of these radio circuits was brought into operation between London and New York in 1927. As demands for service have grown, the number of circuits has been increased. We are, however, fast approaching a limit on further additions, as almost all possible frequency space has now been occupied. The submarine telephone cable has come therefore at an opportune time; further growth in traffic is not limited by traffic capacity.

Technical developments over the years by the British Post Office and Bell Telephone Laboratories have brought continuing improvement in the quality, continuity, and reliability of the radio circuits. The use of high-frequency transmission on a single side band with suppressed carrier and steerable receiving antenna are typical of these developments. Even so, the route, because of its location on the earth's surface, is particularly susceptible to ionospheric disturbances which produce quality deterioration and at times interrupt the service completely. Cable transmission is free of all such quality and continuity limitations. In fact, service of the quality and reliability of the long-distance service in America and Western Europe is possible. This quality and continuity improvement may well accelerate the growth in transatlantic traffic.

The British Post Office and Bell Telephone Laboratories are continuing vigorous programs of research and development on submarine cable systems. Continuing technical advance can be anticipated. Broader transmission bands, lower cost systems, and greater insurance of continuous, reliable, and high-quality services surely follow.

Vol. 7
Pt. 1
1956

Transatlantic Telephone Cable System—Planning and Over-all Performance

E. T. MOTTRAM R. J. HALSEY J. W. EMLING R. G. GRIFFITH

Synopsis: The transatlantic telephone cable system was designed as a link connecting continent-wide communication networks on the two sides of the Atlantic. The technical planning of the system and the objectives set up so that this role would be fulfilled, are the principal subjects of this paper. Typical performance characteristics illustrate the high degree with which the objectives have been realized. Optimum application of the experience of the British Post Office with rigid repeaters and the Bell System with flexible repeaters, together with close cooperation among the administrations, have played a large part in achieving the objectives.

THE transatlantic telephone cable system was planned primarily to connect London to New York and London to Montreal, and thus serve as an interconnection between continent-wide networks on the two sides of the Atlantic. Thus, the system has to be capable of serving as a link in wire circuits as long as 10,000 miles, connecting telephone instruments supplied by various administrations and used by peoples of many nations. This role as an intercontinental link has, therefore, been a controlling consideration in setting the basic objectives for the system.

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957.

E. T. MOTTRAM and J. W. EMLING are with the Bell Telephone Laboratories, Murray Hill, N. J., and New York, N. Y., respectively; R. J. HALSEY is with the British Post Office, London, England; and R. G. GRIFFITH is with the Canadian Overseas Telecommunications Corporation, Montreal, Que., Canada.

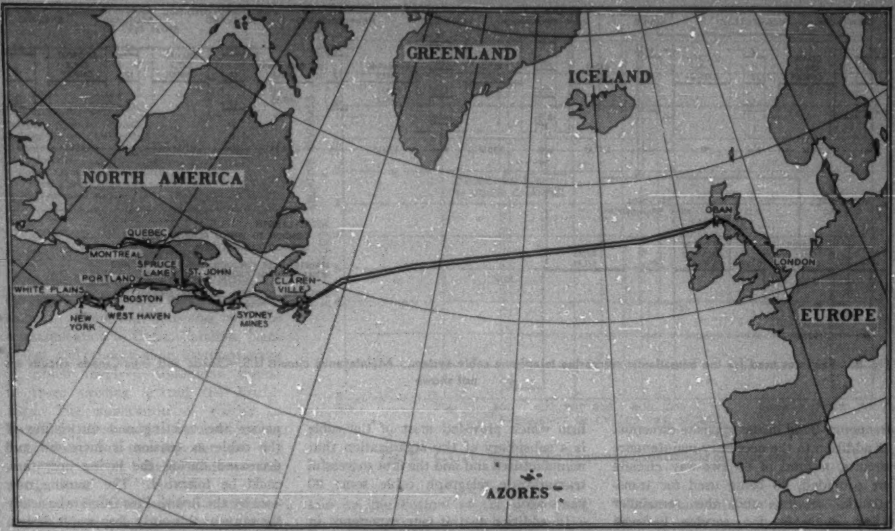


Fig. 1. Route of the transatlantic submarine telephone cable system

The end sections of the system utilize facilities which are integral parts of the internal networks of the United States, Great Britain, and Canada, but the essential new connecting links, extending between Oban, Scotland, and the United States-Canada border, and forming the greater part of the system, were built under an agreement between the joint owners, the American Telephone and Telegraph Company and its subsidiary the Eastern Telephone and Telegraph Company (in Canada), the British Post Office, and the Canadian Overseas Telecommunication Corporation. It is thus the joint effort of three nations.

In planning the system, the main centers of interest were, naturally, the two submarine cable sections, Scotland to Newfoundland, and Newfoundland to Nova Scotia, each of which had to meet a unique combination of requirements imposed by water depth, cable length, and transmitted bandwidth.

Over-All View of the System

The transatlantic system provides 29 telephone circuits between London and New York, 6 telephone circuits between London and Montreal, and a single circuit split between London-New York and London-Montreal; this split circuit is available for telegraph and other narrow-band uses. There are also 24 telephone circuits available for local service between Newfoundland and the mainland of Canada, and there is considerable excess capacity over the radio-relay link that crosses the Maritime Provinces of Canada.

A map of the system is shown in Fig. 1. The facilities used, together with the approximate route distances are shown in Fig. 2. It will be seen that the over-all lengths of the London to New York and London to Montreal circuits are 4,078 and 4,157 statute-miles respectively. Seven of the New York to London circuits are permanently extended to European continental centers; Paris, Frankfurt (two), Amsterdam, Brussels, Copenhagen, and Berne. The longest circuit is thus New York to Copenhagen, 4,948 miles.

Starting at London, which is the switching center for United Kingdom and continental points, 24-circuit carrier cables provide two alternative routes to Glasgow and thence to Oban by a new coaxial cable. Between London and Oban the two routes are fed in parallel at the sending ends, so a change-over can be effected at the receiving ends only. At a later date, an alternative route out of Oban will be provided by a new coaxial cable to Inverness.

From Oban a deep-sea submarine link connects to Clarenville, Newfoundland. This link is, in fact, two parallel submarine cables, one used for east-to-west transmission, the other for transmission in the reverse direction. Each cable is roughly 1,950 nautical-miles in length and lies at depths varying between a few hundred fathoms, on the continental shelf and to about 2,300 fathoms at the deepest point. Each cable incorporates 51 repeaters in flexible housings which compensate for the cable attenuation of about 3,200 db (decibel) at the top frequency of 164 kc (kilocycles per second). These cables carry 36 telephone circuits and maintenance circuits and establish the present maximum capacity of the transatlantic system.

At Clarenville, connection is made with Sydney Mines, Nova Scotia, by a second cable system which goes 63 statute-miles overland to Terrenceville, Newfoundland, and thence about 270 nautical-miles in coastal waters at a depth of about 250 fathoms. Although this system is partly on land, it is basically a submarine system in design, the two portions differing only in the protection of the cable. In this link, the two directions of transmission are carried by the same cable, a low-frequency band being used from west-to-east and a high-

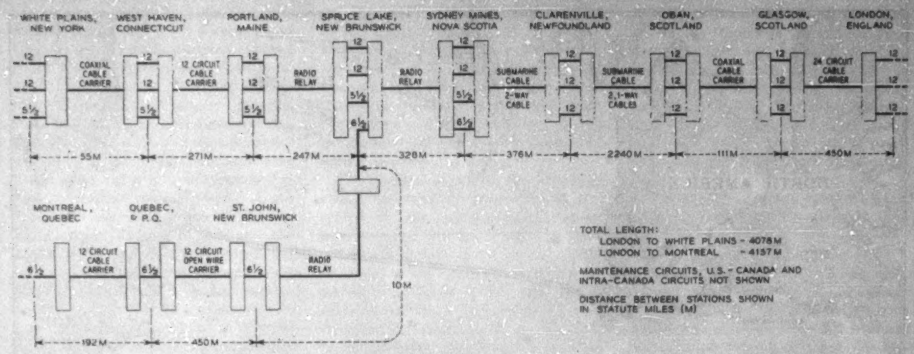

WHITE PLAINS, NEW YORK
WEST HAVEN, CONNECTICUT
PORTLAND, MAINE
SPRUCE LAKE, NEW BRUNSWICK
SYDNEY MINES, NOVA SCOTIA
CLARENVILLE, NEWFOUNDLAND
OBAN, SCOTLAND
GLASGOW, SCOTLAND
LONDON, ENGLAND
12
12
5½
COAXIAL CABLE CARRIER
12 CIRCUIT CABLE CARRIER
RADIO RELAY
RADIO RELAY
6½
SUBMARINE CABLE
2-WAY CABLE
SUBMARINE CABLE
2, 1-WAY CABLES
COAXIAL CABLE CARRIER
24 CIRCUIT CABLE CARRIER
55M
271M
247M
328M
376M
2240M
111M
450M
MONTREAL, QUEBEC
QUEBEC, & P.Q.
ST. JOHN, NEW BRUNSWICK
12 CIRCUIT CABLE CARRIER
12 CIRCUIT OPEN WIRE CARRIER
RADIO RELAY
10M
192M
450M
TOTAL LENGTH:
LONDON TO WHITE PLAINS - 4078M
LONDON TO MONTREAL - 4157M
MAINTENANCE CIRCUITS, U.S. - CANADA AND INTRA-CANADA CIRCUITS NOT SHOWN
DISTANCE BETWEEN STATIONS SHOWN IN STATUTE MILES (M)

Fig. 2. Facilities used for the transatlantic submarine telephone cable system. Maintenance circuits U.S.–Canada and Intra-Canada circuits are not shown

frequency band in the opposite direction. In addition to the necessary maintenance circuits, a total of 60 two-way circuits are provided, 36 being used for transatlantic service, and the remainder being available for service between Newfoundland and the Mainland. Sixteen 2-way repeaters in rigid containers provide almost 1,000-db gain at this system's top frequency of 552 kc.

From Sydney Mines, transmission is by radio relay to the United States-Canada border and thence to Portland, Maine. This system operates at about 4,000 mc (megacycles) and includes 17 intermediate stations. From Portland, standard 12-circuit carrier- and coaxial-cable facilities are used to connect with White Plains, New York, the American switching center 30 miles north of New York City, where connection is made to the Bell System network.

The Montreal circuits leave the radio-relay route at Spruce Lake, a relay station near the border, from which point a short radio spur connects to St. John, New Brunswick, thence to Quebec on a 12-circuit open-wire carrier system and thence to Montreal on a 12-circuit cable carrier system.

Background to the Submarine Cable Systems

The submarine cable sections have been built upon a long background of experience. Some of the cable laying and design techniques go back to the early telegraph cables of almost a century ago, and Lord Kelvin's analysis of the laying process is still the standard mathematical treatise on the subject. It is also interesting to note that the firm which provided most of the cable is a subsidiary of the organization that manufactured and laid the first successful transatlantic telegraph cable some 90 years ago.

In addition to the long experience in submarine telegraphy, the transatlantic system has drawn on over a quarter of a century of experience of telephone cable work in the British Post Office and the Bell System. Experience in these two organizations has been quite different, but each in its own way has been invaluable in achieving today's system.

British Experience

In Great Britain, communication to the continent dominated the early work in submarine telephony and led to systems providing relatively large numbers of circuits over short cables laid in shallow water. Early systems were unrepeatered, but the advantages of submerged repeaters were apparent. Experimental work, started in 1938, culminated in the first submerged repeater installation in an Anglesey-Isle of Man cable in 1943. Currently, there are many repeaters in the various shallow-water cables radiating from the British Isles.

These repeaters, although of a size and mechanical structure well suited to shallow-water applications, are not structurally suited to Atlantic depths. In 1948, the Post Office began to study deep-water problems, and the first laying tests of a deep-water repeater housing were conducted in the Bay of Biscay in 1951. This housing was rigid, like the shallow-water ones, but smaller and double-ended so that the repeater was in line with the cable. Thus the rotation of the repeater, which accompanies the twisting and untwisting of the cable as tension is increased and decreased during the laying operation, could be tolerated. The housing now used by the British Post Office is basically the same as this early deep-water design, although minor modifications have been made to improve the closure and water seals.

A serious study of transatlantic telephony was begun by the Post Office in 1950 when a committee was set up to report on future possibilities of repeatered cables. As a result, it was decided, in 1952, to engineer a new telephone cable to Scandinavia, 300 nautical-miles in length, as a deep-water prototype, even though the requirements of depth, length, and channel capacity all could have been met by existing shallow-water designs.

All of the Post Office submarine systems are alike in that they use but a single cable, the go and return paths being carried by different frequency bands. The adoption of this plan was greatly influenced by the conditions under which the art developed. Because North Sea and Channel cables were highly subject to damage from fishing operations, it was desirable to limit the effects of such damage as much as possible. A single-cable system is obviously preferable under these circumstances to a system using separate go and return cables which could be put out of service by damage to either cable. Since these systems were designed for shallow-water use, the disadvantage of the additional container size required for 2-way repeaters was of no great moment compared to the advantages of a single-cable system.

Table I. Objectives for Standard Deviation of Net Loss

Link	Standard Deviation, Db
New York–Portland	0.75
Portland–Sydney Mines	0.75
Sydney Mines–Clarenville	0.5
Clarenville-Oban	0.5
Oban-London	0.75
Total, assuming rms addition	
New York–London Montreal–London	1.5

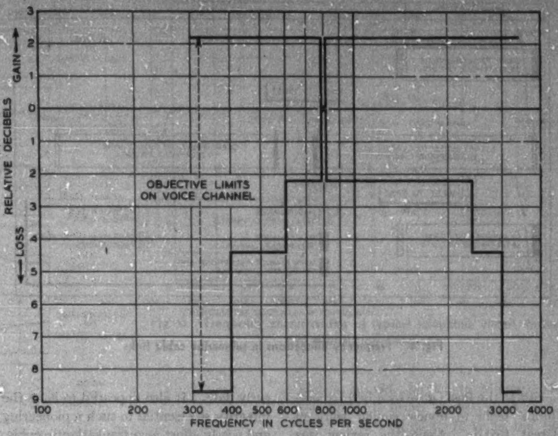


Fig. 3. C.C.I.F. objective for frequency characteristic of voice channel

UNITED STATES EXPERIENCE

In the United States, the cable art developed under very different circumstances. There was, of course, need for communication to Cuba, Catalina, Nantucket, and other off-shore locations, some of which involved conditions similar to those existing around the British Isles. The application of carrier to several of these cables occurred at an early date, but the repeater art was not directed at these shallow-water applications.

For many years, telephone communication to Europe had been an important goal and some 35 years ago a specific proposal was made by the Bell System to the Post Office for a single, continuously loaded, nonrepeatered cable to provide a single telephone circuit across the Atlantic.

This system was never built, partly because of the economic depression of the early 1930's and partly because short-wave radio was able to meet current needs. Cable studies and experiments in the laboratory and field were continued, however, and largely influenced subsequent developments. It was at this time that the physical structure of the cable now used in the transatlantic system was worked out. It was also at this time that the harmful effects of physical irregularities in the cable were demonstrated. As cables are laid in deep water, high tensions are developed which unwrap the armor wires that normally spiral about the central structure. As tension changes during the laying process, twisting and untwisting occurs which is harmless if distributed along the cable. But obstructions in the cable which prevent rotation, or any other process, such as starting and stopping of the ship which tends to localize twisting are likely to cause kinking of cable and buckling of the conductors.

By 1932, electronic technology had advanced to a point where serious consideration could be given to a wide-band system with numerous long-life repeaters laid on the bottom of the ocean and powered by current supplied over the cable from sources on shore.

The hazardous effects of obstructions in the cable demonstrated in early laying tests, indicated that the chances of a successful deep-sea cable would be greatest if the repeaters were in small-diameter flexible housings which could pass through laying gear without stopping the ship and without restricting the normal untwisting and twisting of the cable. The structure ultimately evolved, consisting of two overlapping layers of abutting steel pressure rings within a flexible waterproof container, was an important influence on the electrical design, since it placed severe limitations on size and placement of individual components.

Because these repeaters were to lie without failure for many years on the ocean bottom, it was necessary either to provide a minimum number of components of the utmost reliability, or to provide duplicate components to take over in case of failure. The size limitation favored the former approach. Similarly, the need for small size and minimum number of components militated against the use of 2-way repeaters with their associated directional filters.

Out of these considerations grew the Bell System approach to solving the transatlantic problem by the use of two cables, each with built-in flexible amplifiers containing the minimum number of components of utmost reliability and a life objective of 20 years or better.

It was not until the end of World War II that such a system could be tried. At this time, it was decided to install a pair of cables on the Key West–Havana route to evaluate the transatlantic design which had evolved in the prewar years. After further laying trials, this plan was completed in May 1950, with the laying of two cables. Each of these had three built-in repeaters lying at depths up to 950 fathoms. These cables, each about 120 nautical-miles in length, carry 24 telephone circuits. They have now been in continuous service for over 6 years without repeater failure or evidence of deterioration.

Early Transatlantic Technical Decisions

Early in 1952, negotiations concerning a transatlantic cable were again opened between the American Telephone and Telegraph Company and the British Post Office. As indicated previously, at that time each party had been laying plans for such a system. Thus it became necessary to evaluate the work on each side of the Atlantic to evolve the best technical solution.

To do this, a technical team from the Post Office visited the Bell Telephone Laboratories, in the Fall of 1952, to examine developments in the United States. The work of the preceding 30 years was reviewed in detail with particular emphasis on the development and manufacture of the 1950 Key West–Havana cables. This was followed by

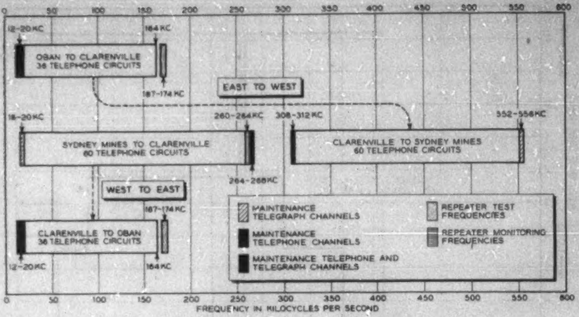


Fig. 4. Frequency allocations in submarine cable links

a visit to the Post Office by a Bell Laboratories' team to review similar work in Great Britain. Again the review was comprehensive, covering shallow-water systems as well as plans for deep-water repeaters. Each visit was characterized by a frankness and complete openness of discussion that is perhaps unusual in international negotiations.

As is apparent from the previous discussion, it was found that the basic features of a deep-water design had been completed by the Bell System. Not only had many of the components been under laboratory test for many years, but a complete system had been operating for $2^1/_2$ years between Havana and Key West. To use a phrase coined at the time, the design had proven integrity.

Because of the years of proof and the conservative approach adopted to assure long life, the design was far from modern. The electron tubes, for example, had characteristics typical of tubes of the late 1930's, when, in fact, they were designed. Similarly, other components were essentially of prewar design.

The Post Office, on the other hand, had pioneered shallow-water repeaters and were pre-eminent in this field. Their deep-water designs were still evolving and had not yet been subjected to the same rigorous tests as the Bell System repeaters. This later evolution, however, made possible a much more modern design. The electron tubes, for example, had a mutual conductance of 6,000 micromhos as compared with about 1,000 in the Bell System repeater, and thus had a potentiality for much greater repeater bandwidths.

It was apparent from these reviews that only the American design was far enough advanced to assure service at an early date. It also appeared to have the integrity so essential to such a pioneering and costly effort as a transatlantic cable. On the other hand, the more modern Post Office design had many elements of potential value. If deep-water laying hazards could be overcome and proof of reliability established, it gave promise of greater flexibility and economy for future systems.

It was on these grounds that Dr. Mervin Kelly, for the Bell System, and Sir Gordon Radley, for the Post Office, jointly recommended that the Bell System design be used for the long length and great depths of the Atlantic crossing, and the Post Office design be used for the Newfoundland–Nova Scotia link where the shallower water afforded less hazard and better observation of this potentially interesting design. The decision to use the Post Office design was subject to technical review after deep-sea laying tests and further experience with circuits and components. This review, made in June 1954, confirmed the soundness of the original recommendation.

System Planning

Planning of the individual systems began as soon as the technical decision just mentioned had been reached. By the time administrative agreements had been reached and the contract signed on Nov. 27, 1953, both parties were ready to set up system objectives and an over-all system plan. This work, too, was accomplished by a series of technical meetings held alternately in the United States and the United Kingdom, with additional meetings in Canada.

At the first of these meetings, a decision of far-reaching importance was made. It was agreed that each technical problem would be solved as it arose in so far as possible on the best engineering basis, putting aside all considerations of national pride. Adherence to this principle did much to forward the technical negotiations.

The initial joint meeting was also responsible for establishing most of the basic performance objectives of the system. The target date for opening of service, Dec. 1, 1956, had been settled even earlier and was, in the event, bettered by nearly 10 weeks.

SERVICE OBJECTIVES

A statement of the manner in which the system would be used and the services to be provided was a necessary preliminary to establishing performance objectives.

It was agreed that the system should be designed as a connecting link between the North American and European long-distance networks. As such, it should be capable of connecting any telephone in North America, ordinarily reached through the Bell System or Canadian long-distance networks, with any telephone in the British Isles or with any telephone normally reached from the British Isles through the European continental network. The system would be designed primarily for message telephone service, but consideration would be given to the provision of other services such as VF carrier telegraph, program (music), and telephotograph, as permitted by technical and contractual considerations. It was also agreed that the two submarine cable links should be so planned that it would be possible to utilize the full bandwidth in any desired manner in the future. Thus, for example, repeater test signals should be outside the main transmission band.

All elements in the submarine cable systems were to be planned for reliable service over a period of at least 20 years.

TRANSMISSION OBJECTIVES

The term objective was used advisedly in describing the aims of the system. It was agreed that such objectives were not

Table II. RMS Noise Objectives in Busy Hour

Link	Approximate Mileage	Noise, Dba
New York–Sydney Mines Montreal–Sydney Mines	1,000	31
Sydney Mines–Clarenville	400	28
Clarenville-Oban	2,000	36
Oban-London	500	28
Total		
New York–London Montreal–London		38

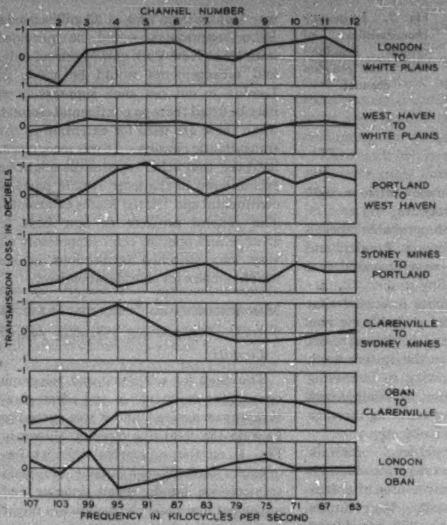


Fig. 5. Frequency characteristics of typical London—New York channel group. (Measured at group frequencies corresponding to 1,000 cycles)

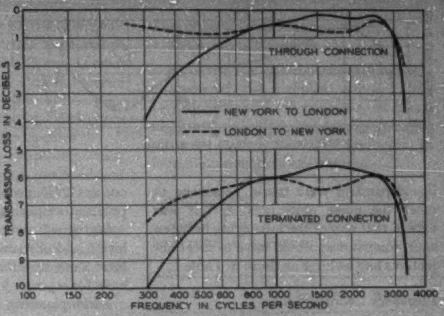


Fig. 6. Frequency characteristics of typical telephone circuit, in the two transmission directions

ironclad requirements but rather desirable goals which it was believed practical to attain with the facilities proposed. Reasonable departure from these goals, however, would not be reason for major redesign.

Since the transatlantic circuits were to connect two extensive networks, the broad objective was to add as little loss and other forms of impairment as practical. To this end, they were to be designed essentially to the standards of international circuits as defined by the CCIF (International Consultative Committee on Telephony)* and of circuits connecting main switching points in national networks, as for example, Regional Centers in the Bell System network and Zone Centers in the Post Office network.

The possibility of increasing the circuit capacity of the system by using channel spacings less than 4 kc was obvious. It was decided, however, to adopt, initially at least, the 4-kc spacing commonly used by long-distance systems on both sides of the Atlantic. This would make possible the use of standard multiplexing arrangements, and it was believed that the number of circuits provided would be adequate for the first few years of operation. It would undoubtedly be desirable to increase the number of circuits in later years, but a decision on the method to be used was left until completion of exploratory work on several methods which promised capacity increases with less degradation than narrow-band operation.

The decision to use standard terminal equipment led naturally to acceptance of the principle that the 36 circuits across the Atlantic would be assembled as three 12-channel groups in the range 60 to 108 kc, and the 60 circuits between Newfoundland and Nova Scotia as five 12-channel groups and thence as a supergroup in the range 312 to 552 kc. These are standard modulation stages in the multiplexing arrangements for broadband carrier system on both sides of the Atlantic. Two of the 12-channel transatlantic groups would be connected to New York, and the third would be split to provide $6^{1}/_{2}$ circuits to Montreal and $5^{1}/_{2}$ to New York, in accordance with the agreement.

To provide for program circuits, three eastbound and three westbound channels in each of the three transatlantic groups would be made available when required. Equipment would be provided to replace either two or three 4-kc message telephone channels by a music channel. In order to avoid the agreed group pilot frequencies and to provide service to Montreal, it was agreed to utilize the frequency bands 68 to 76 kc and 64 to 76 kc in the 12-channel groups for this purpose. Terminals of British Post Office design would be used at all points for translation between program and carrier frequencies. The normal Bell System terminals could not be used since they occupy the frequency ranges 80 to 88 kc and 76 to 88 kc which are not compatible with the split group arrangement or with the 84.08-kc end-to-end pilot.

Net Loss

The nominal 1,000-cycle net loss objective between London and New York for calls switched to other long-distance trunks at each end, i.e., the via net loss, was set at 0.5 db. For calls terminating at either New York or London, the loss would be increased by switching a 3.5-db pad in London, as recommended by the CCIF, and a 2-db pad at New York, as standard in the Bell System. Thus a New York to London call would have a net loss of 6 db.

Variations from these nominal net losses owing to temperature effects, lack of perfect equalization and regulation, etc., are to be expected, and a standard deviation of 1.5 db was set as the objective for such variations in the absence of trouble. The allocation of this variation to the various links is shown in Table I.

It is interesting to note that a smaller variation was allocated to the submarine links than to the overland links. It was believed that the more stable environment on the ocean floor would make it possible to meet the rather small variation assigned to these links.

While these loss variations are consistent with normal long-distance trunk objectives, they would not be satisfactory

---

* The CCIF bases its recommendations on a circuit 2,500 km (kilometer) (1,600 miles) in length, with implied pro rata increases for noise impairment.

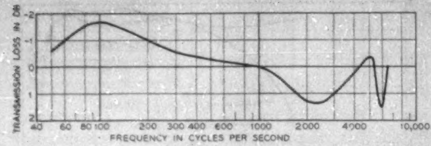


Fig. 7. Frequency characteristics of typical London–New York program channel

if compandors were found necessary to meet the noise objectives, and it was agreed that any of the links lying between such compandors would have to meet objectives half as large as those in Table I.

Frequency Characteristics

For telephone-message circuits, the frequency characteristic recommended by the CCIF, Fig. 3, was adopted with the expectation that it could be bettered by a factor of two, since channel equipments would be included at the circuit terminals only, as described later.

No specific objectives were agreed upon for the frequency characteristics of the 12-channel groups as such, but there was an expectation that ±2 db could be achieved, except for frequencies adjacent to the filters in the split group.

For program channels, the CCIF recommendations were also adopted in respect of the 2-band, 6.4 kc, and 3-band, 10 kc, arrangements. To meet the requirements of these channels and of telegraphy, an over-all frequency-stability objective of ±2 cycles was adopted.

Noise and Crosstalk

Noise objectives were established to be reasonably consistent both with Bell System and CCIF objectives for circuits of transatlantic length. The methods specified by them for the assessment of circuit noise differ in three respects; the units employed, the frequency-weighting employed, and the fraction of the busy hour for which the specified noise may occur. The meters concerned are the Bell System *2B* noise-meter (F1A weighting network) reading in dba (decibels adjusted) and the CCIF psophometer (1951 weighting network) reading in millivolts across 600 ohms. The relationship between readings on the two meters is discussed in a later paper and it will suffice here to note that, for white noise dbm (decibels referred to one milliwatt) (CCIF) = dba (Bell) − 84.

The objective for the rms circuit noise at a zero level point in the busy hour was agreed as 38 dba, i.e., −46 dbm or 3.9 mv. This was allocated between the various links as in Table II.

For the program channels, the agreed noise objective was −50 dbm as measured on a CCIF psophometer with a 1951 program-weighting network.

Statistical data on probable speech levels and distributions at London and New York terminals were provided as a basis for repeater loading studies.

Early planning studies indicated that these objectives would probably be met on all, or nearly all channels without resort to compandors. If, as the system aged, the noise increased owing to increasing misalignment, the use of compandors would offer a means of reducing message circuit noise below the objectives.

The minimum equal-level crosstalk loss between any two telephone channels was set at 56 db for any source of potentially intelligible crosstalk. For channels used for VF telegraph, the equal-level crosstalk loss between go and return directions was set as a minimum of 40 db. For all program channels, the minimum crosstalk attenuation would be 55 db.

Restrictions of Telegraph and Other Services

Since the system was being designed primarily for message telephone service, it was agreed that a channel used for any other service should not contribute more to the system rms or peak load than if this channel were used for message telephone, except by prior agreement between the engineering representatives of the Post Office, Bell System, and Canadian Overseas Telecommunication Corporation.

SIGNALING OBJECTIVES

In order to conserve frequency space, it was decided to transmit all calling and supervisory signals within the telephone channel bands. To avoid transmission degradation, it was agreed that the signaling power and duration would not amount to more than 9 milliwatt-seconds in the busy hour at a zero level point. This would not contribute unduly to the loading of the system.

It was agreed that, for initial operation, ringdown signaling would be employed, but the system design should be such as to permit the use of dialing at a later date.

ECHO SUPPRESSORS

Echo control was considered essential, since the via net loss of the transatlantic circuits would be only 0.5 db, with a one-way transmission time of 35 milliseconds. Echo suppressors would be provided initially at New York and Montreal only, and arrangements would be made in London to cut out such suppressors as may be fitted there on continental circuits, when these are used for extension of the transatlantic circuits. It was recognized, however, that other suppressors might be encountered in the more remote parts of continental and United States extensions. The general problem of how best to arrange and operate echo suppressors on very long switched connections is one which remains for consideration later.

MAINTENANCE AND OPERATING SERVICES

Telephone Speaker and Telegraph Printer Circuits

The need for telephone and telegraph circuits for maintenance and administration was recognized, and it was agreed to provide the following circuits, shown in Fig. 4, on the submarine links at frequencies immediately outside the main transmission bands, where inferior and somewhat uncertain characteristics might be expected: 1. A 4-kc band, possibly substandard in regard to noise, equipped with band-splitting equipment (EB Banks) to provide two half-bandwidth telephone (speaker) circuits, 2. Two frequency-modulated telegraph (printer) circuits. These circuits would be extended over the land circuits to the terminal stations by standard arrangements as needed, and would be used to provide the following facilities:

1. An omnibus speaker circuit connecting the principal stations on the route, including Montreal.
2. A speaker circuit for point-to-point communication between the principal stations, i.e., noncontinuous.
3. A direct printer circuit between London and White Plains.
4. An omnibus printer circuit as 1.

Repeater Test Frequencies

In each submarine cable link, test frequency bands were required for monitoring repeater performance. These are indicated in Fig. 4.

Pilot Frequencies

It was agreed to provide pilot facilities throughout the route for line-up maintenance and regulation purposes. In addition to the usual pilots on the inland networks, there would be provided:

1. A 92-kc pilot in each 12-channel group, continuous only in a particular section of the route and fitted with a recording voltmeter at the receiving end of that section.

Vol.
Pt. 1
1956

2. An 84.08-kc over-all pilot in each 12-channel group as recommended by the CCIF. This would transmit continuously over the entire route and would be monitored and recorded at every main station.

CONNECTIONS BETWEEN COMPONENT LINKS

At the time that the objectives were being established, a far-reaching decision was made to employ channel equipment at London, New York, and Montreal only, and to adopt the frequency band 60 to 108 kc as the standard frequency for connecting the various parts of the over-all system. By adopting this band as standard for the transatlantic system, it also became possible to interconnect readily with land systems at each end.

This agreement also facilitated decisions on responsibility for design and manufacture of equipment. For example, it became logical to define each submarine system as the equipment between points where the 60- to 108-kc band appeared, i.e., the group connecting frames. Thus, these systems would include not only the cable, repeaters, and power supplies, but also the terminal gear to translate between 60 to 108 kc and line frequency of the submarine system. It also became logical to assign responsibility for the manufacture of all of this equipment to the administration responsible for the specific system design, i.e., responsibility for the Oban-Clarenville link to the Bell System and the Clarenville–Sydney Mines link to the Post Office.

The Realization of the System

With decisions reached on the system objectives and interconnecting arrangements, it became possible to lay out jointly a detailed over-all plan and for each administration to proceed with developing and engineering the links under its jurisdiction.

There was an understanding that there should be no deliberate attempt to make the characteristics of one link compensate for those of another, and so it would be incumbent on the administrations to produce the best possible group characteristic on each link.

The over-all plan for the system, as finally developed, is shown in Fig. 2. Except for the necessity to split one of the three transatlantic groups in each direction to provide 6½ circuits to Montreal and 5½ to New York, which required specially designed crystal filters, no unusual circuit facilities were required.

Special equipment arrangements were called for at Sydney Mines and Clarenville to provide security for the Montreal-London circuits where they appeared in the same office with White Plains–London circuits. In these cases, a special locked room was constructed to house the equipment associated with the channel group containing the Canadian circuits.

The details of how the two all-important submarine cable links were designed and engineered to meet their individual objectives are given in companion papers. The efficiency and integrity of these two links are the highest that could be devised by engineers on both sides of the Atlantic.

Finally, each section of the connecting links was lined up and tested individually before bringing them all together as an integrated system.

Over-all Performance of the System

The system went into service on Sept. 25, 1956, so soon after completion of some of the links that it was not possible to include all the final equalizers. Nevertheless, after completion of the initial over-all line-up, the performance has been found to meet very closely the original objectives. The system went into service without the use of compandors on any of the telephone circuits, but compandors are included in the program equipment. At the time of this writing, only the 2-channel program equipment is available for use.

FREQUENCY CHARACTERISTICS OF 12-CHANNEL GROUPS

Fig. 5 shows the frequency characteristic of one of the 12-channel groups, link by link and over-all, measured at group frequencies corresponding to 1,000 cycles on each channel. In both of the complete London-New York groups, the deviation from flat transmission is within ±1.5 db, and some further improvement is to be expected when the equalization is finalized. For the split group, the characteristics are similar except for the effect of the splitting filters.

VARIATION OF OVER-ALL TRANSMISSION LOSS

The system has, of course, only been completed for a short time, but the indications so far are that the standard deviation of the transmission loss, as indicated by the 84.08-kc group pilots is well within the objective of 1.5 db. Alarms operate when the received pilot level deviates by ±4 db and, so far, these alarms have not operated under working conditions.

FREQUENCY CHARACTERISTICS OF TELEPHONE CIRCUITS

Fig. 6 shows the measured frequency characteristic of a typical circuit in the two directions of transmission, as measured in the through and terminated conditions. Half of the CCIF limits are, in fact, met on most of the circuits.

CIRCUIT NOISE

The circuit noise, referred to a zero level point, is as follows:

Link	Best, Dba	Worst, Dba
London–New York	30	36
New York–London	29	41
London–Montreal	30	33
Montreal–London	30	31

Two circuits at present exceed the objective of 38 dba in the New York-London direction only; the higher noise levels refer to the high-frequency channels in the Oban-Clarenville cable. After additional data on the effect of cable temperature variations are accumulated, refinements will be made in the equalization and adjustment of levels on the Oban-Clarenville link. It is expected that the two worst channels can then be made to meet the objectives, still without the use of compandors.

FREQUENCY CHARACTERISTICS OF PROGRAM CHANNELS

Fig. 7 shows the measured frequency characteristic of a London-New York program channel. This is typical.

LONDON-MONTREAL TELEGRAPH CHANNELS

In the agreement, it was envisaged that at least six 50-baud telegraph channels could be provided in each direction in the Canadian half circuit. In fact, 11 such channels have been provided using carriers spaced at 120 cycles and frequency modulation. The telegraph distortion due to the cable system with start-stop signals is about 4 per cent in every case, thus making the circuits suitable for switched connections, without regeneration, up to the same limits as inland systems.

Tests over the system indicate that the channel speed can be raised satisfactorily to 80 bauds on at least ten of the channels. By the adoption of synchronous working, it appears that time-division multiplex systems can be operated on these ten channels to double their capacity at a later date.

Conclusions

The transatlantic cable system has presented unique problems in system planning and design. It has been necessary to design the system to connect the facilities of many countries and to provide for cable communication of unprecedented length. But the stringent design objectives necessary to meet these requirements have not been the only challenge to the designer. It has been necessary to meet these objectives with a system which for over 2,000 miles of its length could not be altered to the slightest extent once it had been placed on the ocean bottom. Except for the adjustments which can be made at the shore terminals of the submarine links, it has not been permissible to make any of the multitude of small design changes, substitutions, and adaptations which are so commonly required in new systems to achieve the design objectives.

The success achieved in meeting the original objectives is a measure of the realism of the early planning as well as the diligence with which the project was carried forward to completion. The accomplishment of getting into commercial service a working system with many complex links 6 weeks after the final splice was dropped overboard, and nearly 10 weeks ahead of schedule, is a further tribute to the close co-operation of the technical people of three nations.

System Design for the North Atlantic Link

H. A. LEWIS R. S. TUCKER G. H. LOVELL J. M. FRASER

Synopsis: Specific aspects of the transatlantic cable system are covered in companion papers. The purpose here is to examine the design and performance of the North Atlantic link, including consideration of factors governing the choice of features, a description of the operational design of the facility, and an outline of those measures available for future application in the event that faults or aging require corrective action.

Description of Link

THAT PORTION of the transatlantic system[1] which connects Newfoundland and Scotland consists of a physical 4-wire, repeatered, undersea link of Bell Laboratories design, with appropriate terminal and power feeding equipment in cable stations at Clarenville and at Oban.

The various elements comprising the link are shown in block form in Fig. 1. The termination points at each end of the link are the group distributing frames (GDF), where the working channels are brought down in three groups to the nominal group frequency band, 60 to 108 kc. At the west end, this point provides the interconnection between the North Atlantic and the Newfoundland–Nova Scotia links. At the east end it is the common point between the North Atlantic link and the standard British long-distance plant over which the circuits are extended to London.

Two separate coaxial cables connect Clarenville with Oban, one handling east-to-west transmission, the other west-to-east. Each is about 1,940 nautical miles long. A total of 102 repeaters are installed in the two cables, at nominal intervals of 37.5 nautical miles. The cables also contain a number of simple undersea equalizers which are needed to bring system performance within the specified objectives.

The working spectrum of each cable extends from 20 to 164 kc, providing for thirty-six 4-kc message channels. Below this band are assigned the telephone (speaker) and telegraph (printer) circuits needed for maintenance and administration of the facility. Above the working band, between 167 and 174 kc, are the crystal frequencies which permit evaluation of the performance of each repeater individually from the shore stations.

The signal complex carried by the cables is derived in the carrier terminals from the signals on the individual voice frequency circuits by conventional frequency division techniques such as are employed in the Ball types *J*, *K* and *L* broad-band carrier systems manufactured by the western Electric Company; see Fig. 2. These signals are applied to the cable through a transmitting amplifier which provides necessary gain and protects the undersea repeaters against harmful overloads. At the incoming end of each cable, a receiving amplifier provides gain and permits level adjustment.

Shore equalizers, next to the transmitting and receiving amplifiers, insert fixed shapes for cable length and level compensation. Adjustable units provide for equalization of the system against seasonal temperature changes on the ocean bed, and some aging.

The power equipment[2] at each cable station includes: 1. regular primary power with Diesel stand-by, 2. rotary machines for driving the cable current supplies, with battery stand-by, 3. battery plants for supplying the carrier terminal bays, and 4. last, but by no means least in complexity, the cable current supplies. These latter furnish regulated direct current to a series loop consisting of the central conductors of the two cables, with their repeaters. A power system ground is provided at the mid-point of the cable current supply at each cable station.

Factors Affecting System Design

General

A repeatered submarine cable system differs from the land-wire type of carrier system in two major respects. First, the cost of repairing a fault, and of the concurrent out-of-service time, is so great as to put an enormous premium on integrity of all the elements in the system and on proper safeguards in the system against shore-end induced faults. Second, once such a system is resting on the sea bottom, it is accessible for adjustment only at its ends. These two restrictions naturally had a profound influence on the design of the North Atlantic link.

Proved Integrity

Assurance of reliability dictated use of elements whose integrity had actually been proved by successful prior experience in the laboratory or in the Bell System. This resulted in adoption of a coaxial cable structure explored by the Laboratories soon after the war, field tested in the Bahamas area in 1948, and applied commercially on the USAF Missile Test Range project[3] in 1953. It also resulted

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957.

H. A. Lewis, R. S. Tucker, and J. M. Fraser are with the Bell Telephone Laboratories, Murray Hill, N. J.; and G. H. Lovell is with the Bell Telephone Laboratories, Winston-Salem, N. C.

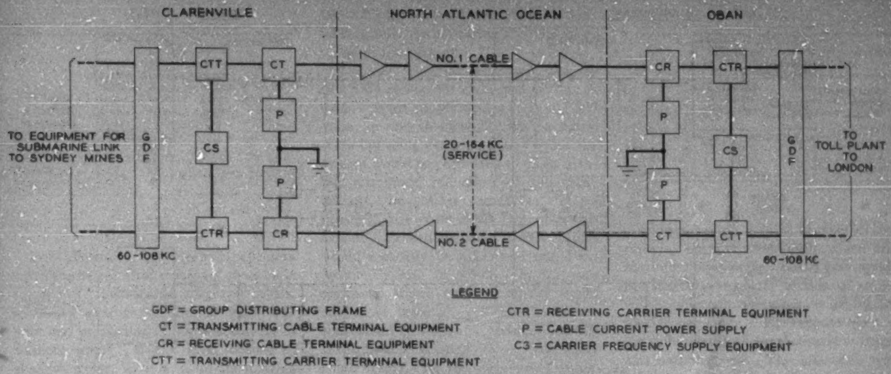

CLARENVILLE
NORTH ATLANTIC OCEAN
OBAN
TO EQUIPMENT FOR SUBMARINE LINK TO SYDNEY MINES
GDF
CTT
CT
CS
P
P
CTR
CR
NO. 1 CABLE
20-164 KC (SERVICE)
NO. 2 CABLE
60-108 KC
CR
CTR
P
CS
P
CT
CTT
GDF
60-108 KC
TO TOLL PLANT TO LONDON
LEGEND
GDF = GROUP DISTRIBUTING FRAME
CT = TRANSMITTING CABLE TERMINAL EQUIPMENT
CR = RECEIVING CABLE TERMINAL EQUIPMENT
CTT = TRANSMITTING CARRIER TERMINAL EQUIPMENT
CTR = RECEIVING CARRIER TERMINAL EQUIPMENT
P = CABLE CURRENT POWER SUPPLY
CS = CARRIER FREQUENCY SUPPLY EQUIPMENT

Fig. 1. Schematic, North Atlantic link

in specification of a basic flexible repeater design which had been under test in the Laboratories since before the war, and had been in use on the Key West–Havana submarine cable telephone system since 1950.[4]

The cable adopted was the largest that had actually been successfully laid in deep water, and its size afforded an important benefit in the form of reduced unit attenuation. The structure and characteristics of this cable are covered in detail in a companion paper.[5] Suffice it to point out here that by its adoption for the North Atlantic link, there were specified for the system designers: 1. the attenuation characteristic of the transmission medium, 2. the influence on this of the pressure and temperature environments, 3. the unit contribution of the cable to the resistance of the power loop, and 4. the impedances faced by the repeaters.

The adaptation[6] of the basic Key West–Havana repeater design to the present project likewise presented the system designers with certain restrictions. Most important, the space and form of the long, tubular structure limited the size of the high-voltage capacitors in the power separation filters, and thus their voltage rating, to the extent that this determined the maximum permissible number of repeaters. Likewise the performance of the repeater circuit was at least partially defined because of several factors. One was the effect of the physical shape on parasitic capacitances in the circuit, which in turn reacted on the feedback and hence on modulation performance, gain-bandwidth, and the aging characteristic. Use of the Key West–Havana electron tube influenced these factors and also tended to fix the input noise figure and load capacity.

In similar respects, the principle of proved integrity reacted into the broad consideration of system design. For instance, in a long system having many repeaters in tandem, automatic gain control (gain regulation) in the repeaters provides an ideal method for minimizing the amount of the total system margin which must be allocated to environmental loss variations. However, this would have required adoption of elements of unproved integrity which might have increased the probability of system failure. So the more simple and reliable alternative was adopted, i.e., fixed gain repeaters with built-in system margins.

System Inaccessibility

The inaccessibility of the undersea system for periodic or seasonal adjustment and the decision to avoid automatic gain regulation were major factors in the allocation of system margins between undersea and shore locations. To avoid wasting such a valuable commodity as margin required the most careful consideration of means of trimming the system during laying. Equalization for control of misalignment in the undersea link is a function of the match between cable attenuation and repeater gain. Generally speaking, these are fixed at the factory. Very small unit deviations from gain and loss objectives could well add to an impressive total in a 3,200-db (decibel) system. Accordingly, it was necessary to plan for periodic adjustment of cable length during laying, and where necessary, insertion of simple mop-up undersea equalizers at the adjustment points.

Design of High-Frequency Line

Terminology

The high-frequency line, as the term is used here, includes the cable, the undersea repeaters, the undersea equalizers, and the shore-station power separation filters, transmitting and line-frequency receiving amplifiers, and associated equalizers.

Repeater Spacing and System Bandwidth

As in most transmission media, the attenuation of the cable increases with increasing frequency. Hence the greater the bandwidth of the system, the greater the number of repeaters needed.

In this system, powered only from its ends, the maximum permissible number of repeaters, and thus the repeater spacing, is determined by a direct voltage limitation, as explained later.

With the repeater spacing fixed, and the type of cable fixed, the required repeater gain versus frequency is known to the degree of accuracy that the cable attenuation is known. The frequency band that can be utilized, then, depends on repeater design considerations, including gain-bandwidth limitations, signal power capacity, and signal-to-noise requirements.

The early studies of the transatlantic system were based on "scaling up" the Havana–Key West system.[4] In these studies consideration was given to extending the band upward as far as possible by using compandors[7] on the top channels, thus lightening the signal-to-noise requirements on these channels by some 15 db, provided they are restricted to message telephone service.

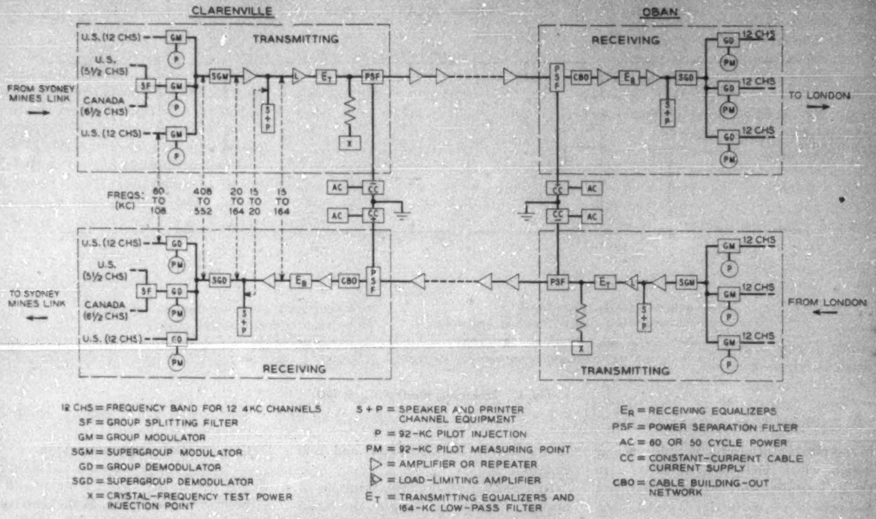

CLARENVILLE
OBAN
TRANSMITTING
RECEIVING
RECEIVING
TRANSMITTING
U.S. (12 CHS)
U.S. (5½ CHS)
CANADA (6½ CHS)
U.S. (12 CHS)
FROM SYDNEY MINES LINK
TO SYDNEY MINES LINK
TO LONDON
FROM LONDON
GM
P
SF
SGM
S+P
E_T
PSF
X
CBO
E_R
SGD
GD
PM
12 CHS
AC
CC
FREQS: (KC)
60 TO 108
408 TO 552
20 TO 164
15 TO 20
15 TO 164
12 CHS = FREQUENCY BAND FOR 12 4KC CHANNELS
SF = GROUP SPLITTING FILTER
GM = GROUP MODULATOR
SGM = SUPERGROUP MODULATOR
GD = GROUP DEMODULATOR
SGD = SUPERGROUP DEMODULATOR
X = CRYSTAL-FREQUENCY TEST POWER INJECTION POINT
S + P = SPEAKER AND PRINTER CHANNEL EQUIPMENT
P = 92-KC PILOT INJECTION
PM = 92-KC PILOT MEASURING POINT
▷ = AMPLIFIER OR REPEATER
▷ = LOAD-LIMITING AMPLIFIER
E_T = TRANSMITTING EQUALIZERS AND 164-KC LOW-PASS FILTER
E_R = RECEIVING EQUALIZERS
PSF = POWER SEPARATION FILTER
AC = 60 OR 50 CYCLE POWER
CC = CONSTANT-CURRENT CABLE CURRENT SUPPLY
CBO = CABLE BUILDING-OUT NETWORK

Fig. 2. Block schematic of terminals, North Atlantic link

As the repeater design was worked out in detail, however, it became evident that a rather sharp upper frequency limit existed. This resulted from the parasitic capacitances imposed by the size and shape of the flexible repeater, the degree of precision required in matching repeater gain to cable loss in such a long system, and the feedback requirements as related to the requirement of at least 20 years' life.

These limitations resulted in the decision to develop a system with 36 channels of 4-kc carrier spacing, utilizing the frequency band from 20 to 164 kc per second.

SIGNAL-TO-NOISE DESIGN

Scaling-up of Key West–Havana System

The length of the North Atlantic link was to be about 16 times that of the Havana–Key West system. The number of channels was to be increased as much as practicable. The length increase entailed an increased power voltage to ground on the end repeaters. Increase in length and in number of channels entailed increased precision in control of variations in cable and repeaters. Work on these and other aspects was carried on concurrently, to determine the basic parameters of the extended system.

Increasing cable size decreases both the attenuation and the d-c resistance and, in turn, the voltage to ground on the end repeaters. It was soon decided that the largest cable size that could be safely adopted was the one used in the Bahamas tests. This has a center-conductor sea-bottom resistance of about 2.38 ohms per nautical mile. It consumes about 28% (per cent) of the total potential drop in cable plus repeaters.

Number of Repeaters

As indicated earlier, the factor which emerged as controlling the number of repeaters was the direct voltage to ground on the end repeaters. Considerations which entered into this were: voltage which blocking capacitors could safely withstand over a life of at least 20 years; voltage which other repeater elements such as connecting tapes between compartments could safely hold without danger of breakdown or corona noise; initial power potential and possible need for increasing direct cable current later in life to combat repeater aging; allowance for repair repeaters; and a reasonable allowance for increased power potential to offset adverse earth potential. Let

R = d-c resistance of center conductor, ohms per nautical mile
L = length of one cable in nautical miles (the length of a cable is greater than the length of the route because of the need to pay out slack. The slack allowance, which averages 5% in deep water on this route, helps to assure that the cable follows the contour of the bottom)
E_{rep} = voltage drop across one repeater at current I
I = ultimate (maximum) line current, amperes
N = ultimate number of submerged repeaters, in terms of equivalent regular repeaters
n = allowance for repair repeaters, in terms of number of regular submerged repeaters using up same voltage drop
E_m = maximum voltage to ground at shore-end repeaters at end of life, and in absence of earth potential
S = spacing of working regular repeaters, nautical miles

Then, for the ultimate condition

$$2E_m = LIR - 2SIR + NE_{\text{rep}} \quad (1)$$

in which the term $2SIR$ accounts for the sum of the cable voltage drops on the two shore-end cable sections; the sum of their lengths is assumed, for simplicity, to equal $2S$. This equation also neglects a small allowance (less than 0.6 volt per mile) for the voltage drop in cable added to the system during repair operations. Also

$$S(N - n + 1) = L \quad (2)$$

because the repeater spacing is determined by the number of working regular repeaters. From equations 1 and 2

$$2E_m = LIR + NE_{\text{rep}} - 2LIR/(N - n + 1) \quad (3)$$

Vol, Pt. 1 195

Table I.

Channel, Frequency	N_{in} = Approximate Input Noise, Dba	G = Approximate Repeater Gain, Db	TL = Approximate Output Trans Level, Db	N_r = Resulting Noise, Dba
Top	−55.5	60.7	− 6	28.4+d_r
Middle	−53	45	−11	20.2+d_r
Bottom	−46	22	−19	12.2+d_r

The allowance *n* for repair repeaters was determined after studies of cable fault records of transoceanic telegraph cables, including average number of faults per year and proportion of faults occurring in shallow water. If the fault occurs in shallow water, as is true in most cases (because of trawler activity, ship anchors, and icebergs) the net length of cable added to the system and the resulting attenuation increase are small. Several shallow-water faults might be permissible without adding a repair repeater. Therefore it was decided to let $n=3$. Since the repair repeater is a 2-tube repeater, while the regular repeater has 3 electron tubes, $n=3$ corresponds to about five repair repeaters per cable.

To determine the maximum voltage E_m, it was necessary to consider blocking capacitors and earth potentials. Based on laboratory life tests, the blocking capacitor developed has an estimated minimum life of 36 years at 2,000 volts. It is estimated that life varies inversely as about the fourth power of the voltage. The potential actually appearing on these capacitors is determined by the distance of the repeater from shore, the power potential applied to the system, and the magnitude and polarity of any earth potential.

Earth potential records on several Western Union submarine telegraph cables were examined. These covered a continuous period from 1938 to 1947, including the very severe magnetic storms of April 1938, and March 1940. It was judged reasonable to allow a margin of 400 volts (200 volts at each shore station) for magnetic storms during the final years of life of the system.

Table II.

Channel, Frequency	M = Estimated Misalignment, Db	d_r = Resulting Noise Increase, Db	N_s = Resulting System Random Noise, Approximate, Dba
Top	12	7.4	36
Middle	10	6.0	26
Bottom	6	3.4	16

With an assumed maximum voltage of 2,300 volts on the end repeaters due to cable-current supply equipment, and 200 volts per end as allowance for the maximum opposing earth potential which the system would be permitted to offset without automatic reduction of the cable current, the voltage across the end repeater in late years of life (i.e., after line current had been increased to offset aging) would normally be 2,300 and would infrequently rise to 2,500. On the rare occasions where earth potential would rise above twice 200 volts, the cable current would be somewhat reduced and the transmission affected to a reasonably small extent. In the early years of life, when the cable-current supply voltage would normally be about 2,000 volts, an opposing earth potential of twice 500 volts could be accommodated without affecting cable current; according to the telegraph cable records this would practically never occur. With conditions changing in this way over the years, the life of the blocking capacitors in the end repeaters was calculated to be satisfactory. Accordingly, E_m was established as 2,300 volts.

The system length *L* was estimated as about 1,985 nautical miles and the ultimate current *I* was estimated as 0.250 ampere, with a corresponding ultimate E_{rep} of about 62.8 volts.

Substituting in equation 3

$N=55$

$N-n=52$ working repeaters

$S=37.4$ nautical miles

From a later estimate of $L=1{,}955$ nautical miles, *S* calculates to be 36.9. The repeater design was based on this spacing in the deep-sea temperature and pressure environment. Subsequently, after better knowledge had been obtained of the cable attenuation in deep water, the actual repeater spacing for the main part of the crossing was changed to about 37.4 nautical miles for the eastbound (no. 1) cable and 37.6 for the westbound. Only 51 repeaters were required in each cable.

Number of Channels

The number of channels which could be transmitted was determined by the upper and bottom boundary frequencies. In this system the bottom frequency was established at 20 kc, primarily because of the loss characteristic of the power separation filters.[6]

Table III.

Channel, Frequency	Weighted Noise in Dba at Zero Transmission Level: Second-Order Products	Third-Order Products	Total
Top	8.2	8.5	11.8
Middle	0.2	3.8	5.4
Bottom	−1.8	2.5	3.9

Preliminary studies were made to estimate the usable top frequency. For a system having a fixed number of repeaters, the frequency falls where the maximum permissible repeater gain equals the loss of a repeater section of cable, which varies approximately as the square root of frequency. The repeater gain is the difference between the repeater input and output levels. The minimum permissible transmission level at the repeater input depends on the random noise (fluctuation noise) contributed by cable and repeaters, and on the specified requirement for random noise. The maximum permissible transmission level at the repeater output may depend on the modulation noise contributed by repeaters below overload, or on overload on the peaks of the multichannel signal complex. In this system, overload was found to be the controlling factor.

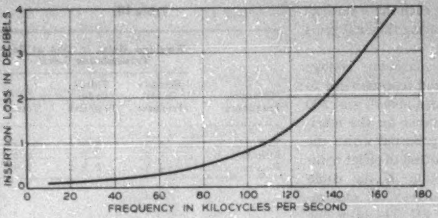
An important consideration was to provide enough feedback in the repeater so that at the end of 20 years the accumulated gain change (Mu-Beta effect) in all of the repeaters would not cause the signal-to-noise performance to fall outside limits. The usable feedback voltage was scaled from the Havana–Key West design according to the relation that this voltage varies inversely as the 5/3 power of the top frequency. Electron-tube aging was estimated from laboratory life tests on the Key West–Havana types of tubes.

Concurrently, detailed theoretical and experimental studies being conducted on the transmission design of a repeater suited to transatlantic use with the chosen type of cable, as discussed in a companion paper.[7] Intimate acquaintance with the

Table IV.

Channel, Frequency	Estimated System RMS Noise at 0.00-Db Transmission Level: Dba	Dbm, Psophometer	Dbm, Psophometer*
Top	36	−48	−48
Middle	26	−58	−58
Bottom	16	−68	−68

* The psophometer is the CCIF circuit noise meter on message telephone circuits. Dbm (psophometer)+84=dba, using CCIF 1951 weighting and Bell System F1A weighting.



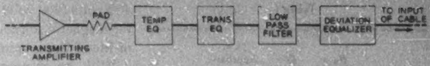


Fig. 3 (left). Undersea equalizer: loss-frequency characteristic

Fig. 4 (above). Transmitting equalizers: block schematic

repeater limitations led to a decision in 1953 to develop a system with a working spectrum of 144 kc (36 channels) and a top frequency of 164 kc. Use of compandors would not appreciably increase this top frequency.

Computed Noise

When the repeater design was established, the remaining theoretical work on signal to noise consisted in refining the determination of the repeater output and input levels; computations of system noise, and comparison of this with the objectives to establish the margin available for variations; and determination of the necessary measures in manufacturing and cable laying so that these margins would not be exceeded by the deviations from ideal conditions which would occur. These deviations assumed great importance, because they tended to accumulate over the entire length of the system, and because many of them were unknown in magnitude before the system was actually laid.

The repeater levels and the resultant system noise were computed as follows: It was recognized that the output levels of different repeaters would differ somewhat at any given frequency. The maximum allowable output level of the highest-level repeater was computed by the criterion that the instantaneous voltage at its output grid should be expected to reach the load-limit voltage very infrequently. This is the system load criterion established by Holbrook and Dixon.[8] It premises that in the busy hour, the load-limit voltage (instantaneous peak value) should be reached 0.001% of the time, or less. It is probable that the level could be raised 2 db higher than the one computed in this way without noticeable effect on intermodulation noise.

An important factor in the Holbrook-Dixon method is the talker volume distribution. Because of the special nature of this long circuit, a careful study was made of the expected United States talker volumes.

First, recent measurements of volumes on long-distance circuits were examined for the relation between talker volume and circuit length. They showed a small increase for the longer-distance circuits. This relation was extrapolated by a small amount to reach the 4,000-mile value appropriate to the New York-London distance.

Second, an estimate was made of the probable trends in the Bell System plant in the next several years, which might affect the United States volumes on transatlantic cable calls.

The result of this was a "most probable US volume distribution." This distribution, which had an average value of −12.5 vu (volume units) at the zero level point, with a standard deviation of 5 db, agreed very well with one furnished by the British Post Office and based on calls between London and the European continent. A further small allowance was then made for the contribution of signaling tones and system pilots, which brought the resultant distribution to an average value of −12 vu at zero level of the system (the level of the outgoing New York or London or Montreal switchboard), and a standard deviation of 5 db. It is approximately a normal-law distribution (expressed in vu), except that the very infrequent high volumes are reduced by load limiting in the inland circuits.

The other data needed for system load computations are the number of channels, and the "circuit activity," i.e., the per cent of time during the busy hour that the circuit is actually carrying voice in a given direction (eastbound or westbound). The circuit activity value used in designing United States long-distance multichannel circuits is 25%; for the transatlantic system 30% was used.

The peak value of the computed system multichannel signal is the same as the peak value of a sine wave having an average power of +17.4 dbm (decibels referred to one milliwatt) at the zero transmission level point of the system.

This value, together with the measured sine-wave load capacity of the undersea repeater, determines the maximum permissible output transmission level of the repeater. The measured sine-wave load capacity is about +13.5 dbm at 164 kc. Hence the maximum permissible output transmission level for the 164-kc channel is $+13.5-17.4=-3.9$ db.

If the relative output levels of the various submarine repeaters were precisely known, the highest-level repeater could have an output transmission level of −3.9 db at 164 kc. An allowance of about 2 db was made for uncertainty in knowledge of repeater levels, giving −6 db as the design value for the maximum repeater output level at 164 kc. (At the time of writing, the no. 1 cable, eastbound, is set up with a somewhat lower maximum output level than this; the safe increase in level will be determined later.)

The repeater has frequency shaping in the circuit between the grid of its output tube and the cable. The maximum repeater output at lower frequencies is smaller than at 164 kc, but the maximum voltage on the output grid, from an overload standpoint, is approximately constant over the 20- to 164-kc band.

The transmission level of the maximum level repeater output is thus determined, based on load considerations. Another factor which might limit this level is modulation noise. This, however, was found to be less of a restriction than the load limitation.

With the output level determined, the random noise and the modulation noise for the system can be computed. The random-noise computation is made on the assumption that all repeaters are at the same level, and then a correction is made for the estimated differences in levels of the various repeaters. The equation is

$$N_0 = N_{in} + G - TL + 10 \log n + d_r$$

where

N_0 = system random noise, in dba (decibels adjusted), referred to zero transmission level. ("Dba" is a term used for describing the interfering effect of noise on a speech channel. Readings of the 2B noise meter with F1A weighting may be converted

IN
OUT
APPROX. LOSS IN DB AT 164 KC

Fig. 5. Transmitting temperature equalizer

to dba by adding 7 db. Dba may be translated to dbm, unweighted, by noting that flat noise having a power of 1 milliwatt over a 3,000-cycle band equals approximately +82 dba, and that 1 milliwatt of 1,000-cycle single frequency power equals +85 dba)

N_{1n} = random noise per repeater in dba, referred to repeater input level

G = repeater gain in db

TL = transmission level of repeater output

n = number of undersea repeaters

d_r = db increase in noise due to differing output levels of the various repeaters as compared to the highest-level repeater

At the top frequency of 164 kc, $TL=-6$ db as seen in the foregoing, $G=60.7$ db, and N_{1n} is about -55.5 dba, which corresponds to a noise figure of about 2 db. Hence, for 52 repeaters, $N_0=-55.5+60.7-(-6)+10\log 52+d_r=28.4+d_r$.

At lower frequencies, the noise power referred to repeater input is greater because part of the equalization loss is in the input circuit; the repeater gain is less, to match the lesser loss of a repeater section of cable; and the repeater output level is lower on account of the equalization loss from output grid to repeater output. This is shown in Table I.

To estimate d_r (which is a function of frequency) before the cable was laid, the factors were studied which might contribute to differences between the levels of the various repeaters. Estimates were made of probable total misalignment, by which is meant the level difference between highest-level and lowest-level repeaters. These values, together with estimates of the resulting noise increases, are shown in Table II. This is based on the assumption that the repeater levels would be distributed approximately uniformly between highest and lowest. The position of a repeater along the cable route is not significant.

A study was made of the expected intermodulation noise. This noise is affected by repeater level differences, by talker volumes, and by circuit activity. It has a time distribution, the most attention being given to the rms modulation noise in the busy hour.

Modulation noise was computed in two ways: by the Bennett method,[9] and by the Brockbank-Wass method.[10] Results by the two methods are in fairly good agreement. The values computed by the Bennett method are shown in Table III.

Because noise powers, not dba, are additive, these values of modulation noise would contribute only a very small amount to the total noise in the deep-sea cable system, as previously stated. The following table for adding two noises expressed in dba shows the magnitudes of the increases.

Db difference between larger and smaller noises	2	4	6	8	10	15	20
Resulting db to be added to larger noise to get sum of the two	2.1	1.4	1.0	0.6	0.4	0.1	0+

Other possible contributory sources of noise in the deep-sea cable system are noise in the terminals, noise picked up in the shore lead-in cables, and corona noise in the repeaters. The system has been designed so that the expected total of these is small. Hence the estimate of total deep-sea cable system noise, made before cable laying, gave the values shown in Table IV, to the nearest decibel.

MISALIGNMENT CONTROL

Objective

Since the noise objective for the deep-sea cable link was set as 36 dba in the joint meetings of the British Post Office and Bell System representatives, the foregoing estimate of system noise led in turn to the objective that the system be manufactured and laid in such a way that the misalignment, including effects of seasonal temperature change, should be no greater in the top channel than the value assumed in the estimate, which was 12 db. Half of this was allotted to initial misalignment and half to effects of temperature change. Greater misalignments were permissible at lower frequencies. Control of this misalignment required extensive consideration in the equalization design of the system.

Causes of Misalignment

The basic causes of misalignment can be grouped as follows: those producing unequal repeater levels when the system is first laid; those resulting from changes in cable loss produced by changes in sea-bottom temperature; and those from aging of the cable or repeaters.

There are a large number of possible causes of initial misalignment. While the cable and repeaters were manufactured within very close tolerances to their design objectives, they could not exactly meet those objectives. In addition, the cable loss as determined in the factory must be translated to the estimated loss on sea bottom, and the length of each repeater section must be tailored in the factory so that its expected sea-bottom loss will best match the repeater gain. Possible sources of error in this process include: uncertainty in temperature of the cable when it is measured in the factory, and in its temperature on the sea bottom; uncertainty in temperature and pressure coefficients of attenuation; changes in cable loss between factory and sea-bottom conditions, not accounted for by pressures and temperature coefficients.

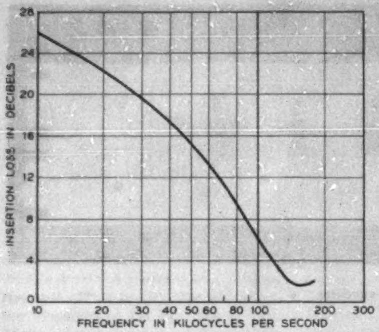


Fig. 6. Transmitting equalizer loss-frequency characteristic

These latter changes in cable loss were called "laying effect." The determination of the magnitude of laying effect, and its causes, are discussed in the paper on cable design.[5] Suffice it to say here that after measuring the loss of a length of cable in the factory and computing the sea-bottom loss, the computed result was a little greater than the actual sea-bottom loss. The difference, i.e., the laying effect, was approximately proportional to frequency, and was greater for deep-sea than for shallow-water conditions. Its existence was confirmed by precise measurements on trial lengths of cable, taken early in 1955 in connection with cable-laying tests made near Gibraltar.

Laying effect had been suspected from statistical analysis of less precise tests of repeater sections of cable generally similar to transatlantic cable, as measured in the factory and as laid in the vicinity of the Bahamas. However, the repeater design had of necessity been established before the Gibraltar test results were known. The consequence was a small systematic excess of repeater gain over computed cable loss, approximately proportional to frequency, and amounting at the top frequency to about 0.04 db per nautical mile, on the average, for deep-sea conditions, and about 0.025 db per mile for shallow-water conditions. While this difference appears small, it would accumulate in a transatlantic crossing of some 1,600 miles of deep sea and 350 miles of relatively shallow sea, to about 75 db at the top frequency. This would be enough to render almost half of the channels useless if no remedial action were taken.

After a system is laid, changes in cable loss or repeater gain may occur. These may be caused by temperature effects and by aging of cable or repeaters.

Comprehensive studies of the expected amounts of temperature change were made both before and after the 1955 laying. (Factual data on deep sea-bottom temperatures are elusive. Many of the existing data were acquired by unknown methods under unspecified circumstances, with apparatus of unstated accuracies. Statistical analysis of selected portions of the data leads to the quoted estimates.) These gave an estimate of 0 to ±1 F (degrees Fahrenheit) annual variation in sea-bottom temperature in the deep-sea part of the route (about 1,600 nautical miles) and perhaps ±5 F on the continental shelves (about 350 nautical miles). Use of these figures leads to a ±5d-b variation in system net loss at 164 kc from annual temperature changes. If the deep-sea bottom temperature did not change at all, the estimated net loss variation due to temperature would be about half of this.

Control of Misalignment

In the control of misalignment, the first line of defense against variations leading to this condition is the design and production of complementary repeaters and cable. This is discussed in companion papers.

Signal-to-noise changes from undersea temperature variations were minimized by providing adjustable temperature equalizers at both the transmitting and the receiving terminals, and devising a suitable method of choosing when and how to adjust them.

Partial compensation for laying effect was carried out at the cable factory by slightly lengthening the individual repeater sections. In the 1955 cable, where the factory compensation was based on early data, the increased loss compensated for about two fifths of the Gibraltar laying effect at the top frequency. For the 1956 cable, the 1955 transatlantic experience was benefited by, and the compensation was increased to nearly twice this amount. Since the loss of the added cable is approximately proportional to the square root of frequency and the laying effect is approximately directly proportional to frequency, the proportion of laying effect compensated varied with frequency, and a residual remained which had a loss deficiency rising sharply in the upper part of the transmitted band.

The remainder of the laying effect which was not compensated for in the factory, as well as other initial variations, were largely compensated by measures taken during cable laying at intervals of 150 to 200 miles. The whole length of the cable was divided into eleven "ocean blocks," each either four or five repeater sections long. At each junction between successive ocean blocks, means were provided to compensate approximately for the excess gain or loss which had accumulated up to that point.

These means were twofold. The first means was adjustment of the length of the repeater section containing the junction. For this purpose, the beginning of each block, except the first, was manufactured with a small excess length of cable. This was to be cut to the desired length as determined by shipboard transmission measurements.

The second means was a set of undersea equalizers. These equalizers were fixed series networks, encased in housings similar to the repeater housings but shorter.

Before the nature of the laying effect was known, it had been planned to have equalizers with perhaps several shapes of loss versus frequency; but to combat laying effect, nearly all were finally made with a loss curve sloping sharply upward at the higher frequencies, as shown in Fig. 3. (This characteristic was based on a statistical analysis of the data on similar cable laid for the Air Force project.) Because the equalizer components had to be manufactured many months in advance and then sealed into the housings, last-minute designs of undersea equalizers were not practical. The proven-integrity principle prevented use of adjustable units. Because the equalizers were series-type, each block junction has located at approximately mid-repeater section to minimize the effects of reflections between equalizer and cable impedances.

The actual adjustments at block junctions were determined by a series of transmission measurements during laying, as described in the companion paper on cable laying.[11] Six equalizers were used at block junctions in the 1955 (no. 1) cable, and eight in the 1956 (no. 2) cable.

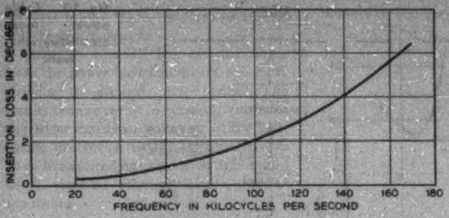


Fig. 7 (above). Transmitting deviation equalizer: loss-frequency characteristic

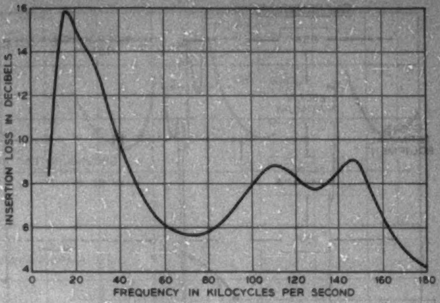


Fig. 9 (right). Receiving fixed equalizer: loss-frequency characteristic

The result of all the precautions taken to control initial misalignment was, in the 1955 cable, to hold the initial level difference between highest-level and lowest-level repeaters to about 6 db near the top frequency and to values between 4 and 9 db at lower frequencies, the 9-db value occurring in the range 50 to 70 kc where there is noise margin. In the 1956 cable, the level difference was about 4 db at the top frequency, and from 2 to 7 db elsewhere in the band.

SHORE EQUALIZATION

The equalization to be provided at the transmitting and receiving ends of the North Atlantic link had these primary functions:

1. For signal-to-noise reasons, to provide a signal level approximately flat with frequency on the grid of the third tube of the undersea repeaters.
2. To equalize the system so that the received signal level is approximately constant over the transmitted band.
3. To keep the system net loss flat, regardless of temperature variations in the ocean. (A change of 1 F in sea-bottom temperature would cause the cable loss to change by 2.8 db at 160 kc, and less than this at lower frequencies. The amplifiers are relatively unaffected by small temperature changes.)
4. To provide overload protection for the highest-level repeater.
5. To incorporate some adjustment against possible cable aging.

A portion of the transmitting terminal is shown in block schematic form in Fig. 4. All of the equalizers are constant resistance, 135-ohm unbalanced bridged-T structures. Their functions follow:

Low-pass Filter

The low-pass filter transmits signals up to 164 kc, and suppresses by 25 db or more, signals in the frequency range from 165 to 175 kc. Its purpose is to prevent an accidentally applied test tone from overloading the deep-sea repeaters should such a tone coincide with the resonance of the crystal used in a repeater for performance checking. The gain of a repeater to a signal applied at the resonance frequency of its crystal is about 25 db greater at 164 kc.

Transmitting Temperature Equalizer

It was estimated that the cable loss change from temperature variations would not exceed ±5 db at the top of the transmitted band. The change in cable loss versus frequency due to temperature variations is approximately the same as if the cable were made slightly longer or shorter, and so the temperature equalizer was designed to match cable shape. Temperature equalizers are used both in the transmitting and receiving terminals to minimize the signal-to-noise degradation caused by temperatures misalignment. Each equalizer provides a range of about ±5 db at the top of the band, the loss being adjustable in steps of 0.5 db by means of keys; see Fig. 5.

This range might appear to be more than necessary but it must be borne in mind that the sea-bottom temperature data left much to be desired in precise knowledge of both average and range.

Transmitting Equalizer

The transmitting equalizer was so designed that, with the temperature equalizer set at mid-range and with repeaters that match the cable loss, the signal level on the grid of the output tube of the first repeater would be flat with frequency. The loss-frequency characteristic is shown in Fig. 6.

Deviation Equalizer

The transmitting deviation equalizer is used to protect the highest-level repeater from overload. The design is based on data obtained during the laying, indicating which repeater was at the highest level at the various frequencies. The characteristic of this equalizer is shown in Fig. 7.

A portion of the receiving terminal is shown in block schematic form on Fig. 8. The generalized function of the equalizers is to make the signal level flat versus frequency at the input to the receiving amplifier no. 2.

The specific functions of the various equalizers are described in the following paragraphs.

Cable Length Equalizer

In planning the system, an estimate was made of how far the last repeater (receiving end) would be from the shore. Considerations of interference and level dictated a maximum distance not exceeding approximately 32 miles. For protection against wave action, trawlers, and similar hazards, the repeater should be no closer than about 5 miles. To take care of this variation, a cable length equalizer was designed that is capable of simulating the loss of 10 miles of cable, adjustable in 0.5-db steps at the top of the frequency band. Two of these could be used if needed. Once a system is laid, this equalizer should require no further adjustment unless it is used to take care of cable aging or a cable repair near shore.

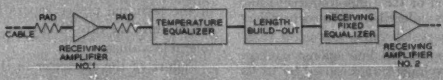

CABLE
PAD
RECEIVING AMPLIFIER NO. 1
PAD
TEMPERATURE EQUALIZER
LENGTH BUILD-OUT
RECEIVING FIXED EQUALIZER
RECEIVING AMPLIFIER NO. 2

Fig. 8. Receiving equalizers: block schematic

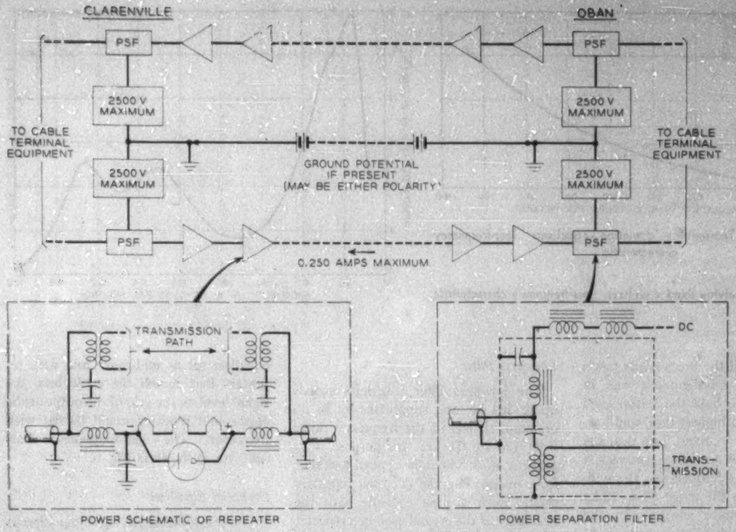


Fig. 10. Cable power loop: elementary schematic

Receiving Fixed Equalizer

The receiving fixed equalizer is the mop-up equalizer for the system. A final receiving equalizer, Fig. 9, has been constructed for the first crossing. Another, tailored to the no. 2 cable, will be designed on the basis of data taken after completion.

Receiving Temperature Equalizer

The receiving temperature equalizer is identical with the transmitting temperature unit.

Operational Design

General

The operational design of a transmission system considers the supplementary facilities which are needed for operation of the main transmission facility, for its supervision, and for its maintenance. In the present instance, these facilities include the cable station power plants for driving the carrier terminals and high-frequency line; the carrier terminals themselves and their associated carrier supply bays; the telephone and telegraph (speaker and printer) equipments needed for maintenance, supervision and administration of the over-all facility; the pilots, protection devices and alarms; and the maintenance and fault-locating equipment.

Power Supplies

With the exception of the plants for cable current supply, the power plants at Clarenville and Oban are relatively conventional, and follow telephone office techniques which are standard for the telephone administration of the particular side of the Atlantic on which they are located. They will not be discussed here, although it might be well to point out that, for the Bell equipment the supply is direct current, obtained from floated storage batteries, while at Oban, the supply for the Post Office equipment is alternating current from rotary machines driven normally from the station a-c supply, with storage battery backup.

The cable current supplies for the North Atlantic link[2] are complex and highly special. It is pertinent to discuss here briefly the requirements which beget this complexity. To assist in this, an elementary schematic of the cable power loop is shown in Fig. 10.

The main requirements that governed the design of these plants are:

1. Constant maintenance of uniform and known current in the power loop.
2. Protection of HF (high-frequency) line against faults induced by failures in power bays.
3. Protection of HF line against damage from power voltage surges caused by faults in the line itself.

Maintenance of constant and known operating current in the line is very important from the standpoint of system life because of the dependence of the rate of electron tube aging on the power dissipated in the heaters.[12]

The principal factor which tends to cause variations in the line current is the earth potential, which may appear between the terminals of the system. This potential may be of varying magnitude and of either polarity. Consequently, it may either aid or oppose the driving potential applied to a particular cable. The power circuit regulation is such that the presence of an aiding potential will be completely compensated, while opposing potentials will be compensated only to the degree possible with the maximum of 2,500 volts applied by the terminal plants. Beyond this point, the line current will be allowed to drop, so as to avoid excessive potential across the power separation filter capacitors in the shore-end repeaters.

Line faults caused by failures or misoperation of the power supplies must be carefully guarded against. Such failures might result in surges on the line, or excessive voltage or current. Protection against surges takes the form of inductors in the terminal power separation filters, which limit the rate of current change to tolerable values.

Failure of the capacitor in these filters could also create dangerous surges. The

Vol. 7
Pt. 1
1956

protection here takes the form of a large voltage design margin.

Faults in the HF line of importance from the power feed standpoint are short circuits to ground, and opens. The former tend to result in excessive currents; the latter excessive voltages. Very fast-acting protection in the terminal power bays is required to cope with these, and series electron tube regulators provide the means.

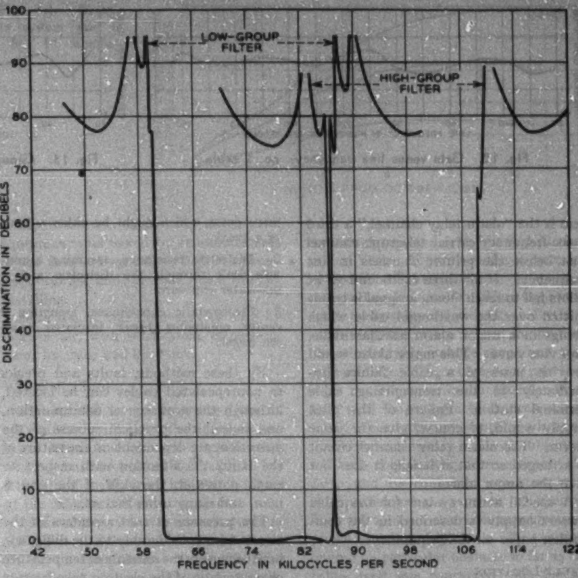


Fig. 11. Splitting filters: loss-frequency characteristic

TERMINAL PLAN

The plan of design of transmission terminal equipment, shown in Fig. 2 was based on the following objectives:

1. Use of standard, or modified standard equipment as far as possible.
2. To facilitate supply and maintenance of standard equipment, use of Bell System equipment in North America and Post Office equipment in the United Kingdom.
3. Provision of full duplicate equipment and means for quick shift between regular and alternate.
4. Provision of special equipment to fit Canadian requirements.
5. Provision of ample order-wire (speaker and printer) equipment, partly because there is no alternate undersea route.
6. Provision of no automatic loss regulation, because the loss changes are so slow; provision of three pilots as a basis for manual regulation.

Much of the equipment is standard. This includes group modems, group connectors, pilot supply, frequency-shift telegraph equipment for printer circuits, etc. Supergroup modulators are modified standard coaxial carrier equipment. The channel modems for speaker and printer circuits are a combination of type-*C* open-wire carrier active equipment, and filters designed for a military carrier system. The transmitting output amplifier is a modified version of one used in the Havana-Key West submarine cable system: the modifications include broader frequency band, lower modulation, and sharper overload cutoff. As in the coaxial carrier system, hybrid coils are provided at points in the Bell equipment where quick patching is needed.

The agreed Canadian bandwidth quota is 26 kc,[1] corresponding to $6\frac{1}{2}$ message telephone channels. The split between the $6\frac{1}{2}$ Canadian channels and the remaining $5\frac{1}{2}$ U. S. channels in the "split group" was accomplished by specially designed group-frequency splitting filters, with very sharp cutoff. The loss-versus-frequency characteristics of these filters are shown in Fig. 11. The resulting Canadian-half telephone channel next to the cut-apart was broad enough to accommodate 12 frequency-shift telegraph channels with 120-cycle carrier telegraph spacing. It is planned to use one of these for automatic regulation of the other 11. The U. S.-half channel is at present unassigned.

Two telephone order wires (speakers) and two teletypewriter order wires (printers) are provided for maintenance and administration purposes.

The two speakers occupy a 4-kc telephone channel which is the lower side band of 20 kc on the line. The 4-kc band is divided into halves by Bell type-*EB* split-channel equipment, which yields two narrow-band telephone channels. One of these is reserved for a "local" Clarenville-Oban circuit, available at all times to the personnel at these terminals. It constitutes a cleared channel which could be of great benefit in emergencies which might arise. The other narrow-band telephone channel, the "omnibus speaker," is extended through to other control points including the system metropolitan terminals.

The two printer circuits are frequency-shift voice-frequency telegraph channels, occupying line frequencies just below 16 kc. These are both brought to direct current at Clarenville and Oban, and extended by carrier telegraph to the metropolitan terminals. One is normally a "through" circuit with teletypewriters at the metropolitan terminals only, and the other is an "omnibus" circuit with teletypewriters connected at Oban, Clarenville, and other control points, permitting message interchange between all connected points.

Ninety-two-kc group frequency pilots are transmitted over the Clarenville-Oban link only, and blocked at its ends. The line frequencies corresponding to the 92-kc group frequencies are 52, 100, and 148 kc. These pilots are used for manual regulation and maintenance of the Clarenville-Oban link. Information derived from them governs the manual setting of the temperature equalizers. The sending power of the 92-kc pilots is regulated to within ±0.1-db variation with time.

Group frequency pilots of 84.080 kc also appear on this link. These, however, are system pilots applied at the metropolitan terminals. They are useful for evaluation of over-all system performance and for quickly locating transmission troubles.

Alarms are provided in the manner usual for multichannel telephone systems. A special feature of the pilot alarm sys-

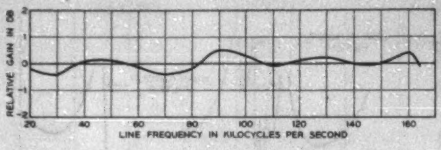


Fig. 12. Gain versus line frequency: no. 1 cable

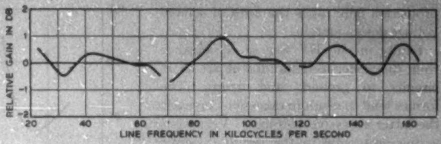


Fig. 13. Group-to-group frequency response: no. 1 cable

tem is the "alarm relay channel," a third voice-frequency carrier telegraph channel just below the printer channels in line frequency. If all three eastbound 92-kc pilots fail to reach Oban, a signal is transmitted over the westbound cable which brings in a major alarm at Clarenville, and vice versa. This major alarm would provide news of a cable failure immediately to the transmitting cable terminal station. Failure of the pilot supply would, of course, give the same alarm. The alarm relay channel circuit is arranged so that if it fails it does not give the major alarm.

A special alarm system for the cable current supply is described in the companion paper on that equipment.[3]

FAULT LOCATION

Length and inaccessibility have always imposed difficult requirements on fault-locating techniques for undersea cable systems. Before the advent of repeatered systems, much effort was expended on use of impedance methods, i.e., those involving resistance and electrostatic capacitance measurements to locate a fault to ground, or a break. (In the literature, "fault location" is a generalized term encompassing the field. A "fault" is an exposure of the cable conductor to the sea without a break in the conductor. A "break" is an interruption of the conductor with or without exposure to the sea.) Some work has been done also on magnetic pickup devices towed along the route by a ship.

With the advent of undersea repeaters, the problem has become even more difficult, both because of the effect of the repeater elements on impedance measurements, and because a fault may interrupt transmission without affecting the d-c loop. It has been necessary therefore to reassess the whole problem of fault location.

LOCATION OF FAULTS AFFECTING THE D-C POWER LOOP

Three types of measurement are made on submarine cables at present:

1. Center conductor resistance, assuming a short circuit or a sea-water exposure at some point, which might be either a fault or a break.
2. Dielectric resistance, assuming a pinhole leak through the dielectric and a sea-water exposure.
3. Electrostatic capacitance, assuming a center conductor break insulated from sea water.

By these methods, faults and breaks in nonrepeatered cables can be located, although the accuracy of determination, and indeed the practical success of the operation, are dependent on the nature of the fault, the situation with respect to earth potential, the skill of the craftsman, and many other factors.

The presence of cold repeaters in the circuit adds considerably to the difficulty, both because of the resistance/temperature characteristic of the electron tube heaters (153 in series with each cable) and because of polarization effects exhibited by the castor oil capacitors in the transatlantic repeaters. The capacitors have a storage characteristic which varies with the magnitude of applied voltage, the duration of its application, and the temperature.

Because of this, the usual methods of fault location are expected to give results of doubtful accuracy and so a new approach is being made to the problem. The results of the work are beyond the scope of this paper.

LOCATION OF TRANSMISSION FAULTS

The previous section deals with the situation where the d-c power loop has been opened or disturbed. Of equal importance is the location of transmission faults when the power loop is intact.

For this purpose, use is made of the discrete frequency crystal provided in each repeater.[6] This crystal is effectively in parallel with the feedback circuit of the amplifier, and at its resonant frequency the amplifier has a noise peak of the order of 25 db, with an effective noise band of about 4 cycles. When one repeater fails, it is possible to recognize the noise peak of each amplifier from the receiving end back to the failed unit. The noise peaks from the faulty unit and all preceding amplifiers will be missing, indicating location of the trouble. Noisy amplifiers can be singled out by an extension of the technique.

TESTING AND MAINTENANCE

Routine testing and maintenance activities at Oban and Clarenville can be divided into four parts: for the carrier terminals, for the station power equipment, for the cable current power units, and for the HF line. Usual methods apply, and the usual types of existing test equipment were provided with one exception.

The exception is a newly designed transmission measuring system. The system employs a decade-type sending oscillator with continuous tuning over the final 1-kc range. Provision is built in for precise calibration of output level. The receiving console contains a selective detector functioning as a terminating meter of 75-ohm input impedance, and can measure over the range −120 dbm to 0.00 dbm. Repeating coils are furnished to permit measurement on 135-ohm circuits. The detector can be calibrated for direct reading when used as a terminating meter.

The system covers a frequency range of 10 kc to 1.1 megacycles. It is useful therefore for measuring much of the standard British- and U. S.-designed carrier terminal equipment, as well as for normal line transmission measurements over both the North Atlantic and Cabot Strait (Newfoundland–Nova Scotia) submarine lines.

A special transmission measuring set is provided in the receiving console which facilitates measurements in the crystal frequency region 166 to 175 kc. As an oscillator, it is capable of delivering an output of up to +8 dbm into a 75-ohm load, with exceptional frequency stability and finely adjustable, motor-driven tuning. This is useful in locating and measuring the narrow-band response peaks of individual repeater crystals. The oscillator frequency is varied in the region of a particular peak, and the received power is measured at the peak frequency and at near-by frequencies. At the peak frequency, the crystal removes nearly all feedback in the repeater. Changes in repeater internal gain can thus be determined from shore.

As a detector, the set can measure from −110 dbm to −60 dbm in a bandwidth of about 2 cycles. This enables it to measure crystal noise peaks which may be spaced as closely as 50 cycles apart. To reduce the random variations in such narrow-band noise, a "slow integrate" circuit, of the order of 10 seconds, is provided. Thus the crystal noise peaks can be compared in magnitude with the system noise level at closely adjacent frequencies.

Assembly and Test of System

General

Assembly and initial testing of the North Atlantic link occupied a span of about 3 years. Because of the geographical and political factors involved, the job was a difficult one and it required close co-operation among individuals in many different organizations on both sides of the ocean.

Clarenville

The cable station at Clarenville houses the carrier terminal, cable-terminating, and power equipment at the junction of the North Atlantic and Cabot Strait links. It was designed by U. S. architects from requirements furnished by American Telephone and Telegraph Company engineers and was approved by the other parties to the enterprise. The building was constructed by a Canadian firm under the supervision of a Canadian architect. The equipment was put in place and connected by installers of the Northern Electric Company and tested by representatives of Bell Telephone Laboratories, Eastern Telephone and Telegraph Company, Northern Electric Company, and the British Post Office.

As Newfoundland is an island, shipments of equipment and supplies to Clarenville were carried by sea or air to St. John's, where they were trans-shipped by rail to Clarenville.

Oban

The cable station at Oban was executed by British contractors from designs drawn up by the British Post Office. Its equipment was installed by a British electrical contracting firm. Two Western Electric Company installers were present at Oban during the installation of the American-made cable terminating equipment and of the cable current supply bays, to provide necessary liaison and interpretation of drawing requirements. The Oban equipment was tested by representatives of the British Post Office, Bell Telephone Laboratories, and the firm which did the installing.

Here too, it was necessary to ship by boat or air, with subsequent trans-shipment by train and by truck.

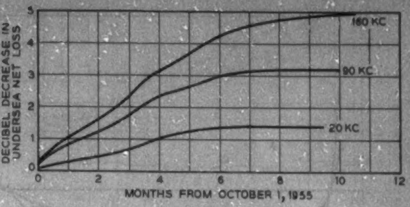


Fig. 14. Change in system gain in first 11 months

Undersea Link

Perhaps the most interesting phase of all was the handling of the undersea section. The actual laying is described elsewhere,[11] but much effort was required before the cable ship ever left the dolphins at the cable manufacturing plants.

Most of the cable was manufactured at the plant of Submarine Cables, Ltd., at Erith, on the Thames about 15 miles downstream from London. A smaller quantity of cable was manufactured by Simplex Wire and Cable Corporation at its plant in Newington, N. H. As the cable was completed, it was coiled in repeater section lengths in huge tanks. The ends of each repeater section were left available at a "splicing platform" where the repeaters, manufactured by Western Electric Company at Hillside, N. J.,[12] were spliced in. Spliced repeaters were located in water-filled troughs for protection against overheating while testing.

All repeaters were armored by Simplex at Newington. This involved their transportation by truck from the Western Electric shop at Hillside, where the flexible repeaters were manufactured, to Newington. While this sounds simple, it was actually a very carefully planned and controlled operation because of the need to avoid subjecting the units to any but the most necessary and unavoidable hazards. Consequently, the truck used for the purpose was specifically selected for size and construction, and was provided with a heating unit in the body. The route was carefully surveyed in advance for unusual hazards, and the truck speed was limited to a very modest value.

After the repeaters were armored, those to be incorporated in Simplex-made cable were spliced at Newington. Those destined for application in British-made cable were shipped by truck, following the same precautions, to Idlewild Airport in New York. From there they proceeded by special freight plane, one or two at a time, to London Airport. Here they were again trans-shipped by truck, to Erith.

The repeaters were housed in shipping cases[13] of a very strange shape and considerable size. These cases were provided with maximum-minimum thermometers and with impactograph devices which would record the maximum acceleration to which the repeaters had been subjected.

When ship-loading time arrived, the cable and its contained repeaters were transported over a system of sheaves from the tanks of the cable manufacturer to the tanks on shipboard. This process offered no particular obstacle so far as the cable itself was concerned, but the requirement against unnecessary bending of the repeater meant that a great deal of special attention had to be given to avoiding unnecessary deviations from a straight run, and to lifting the repeaters around sheaves where the direction of the loading line changed materially. Auxiliary protection was provided on each repeater in the form of angle irons with flexible extensions on each end which served to restrict any bending to the core tube region of the repeater, and safely limited the magnitude of the bends in this region.[11] One further precaution was observed, that of energizing the repeaters during the loading process. This accomplished two things. First, it permitted continuous testing; second, it reduced the hazards of possible damage to the electron tubes during loading, as the tungsten heaters are much more ductile when hot and the glassware of the tube is more resistant to shock and vibration.

System Performance

General

At the time of writing, the no. 2 cable has just been laid, but the no. 1 cable has had 11 months of successful life under test, a preservice period of probably greater duration than has been granted to any land system of comparable length and cost. It has been under constant ob-

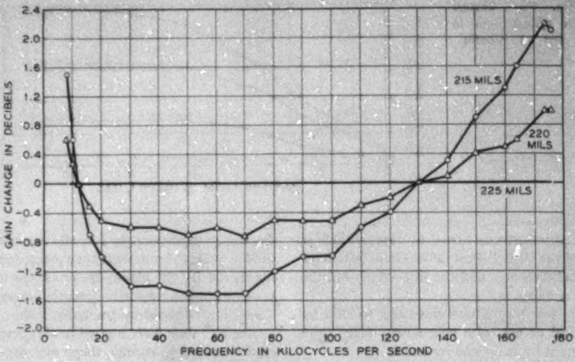


Fig. 15. Effect of cable current on system gain: no. 1 cable

servation and test, largely by the people who will operate it when it goes into service. The preservice measurements have been much more extensive in quantity and scope than the routine tests which are being made now that the system is in service.

Net Loss Tests

Net Loss Versus Frequency

Results of measurements in June 1956 on the no. 1 cable are shown in Figs. 12 and 13. The former covers the equalized HF line. Fig. 13 gives the corresponding group-to-group frequency response of the deep-sea link. These frequency characteristics will vary a little from time to time, depending partly on the change since the last adjustment of the temperature equalizer.

Note that while the slope in gross cable loss between 20 and 164 kc is about 2,100 db, the net loss of the equalized HF line over this band varies only about one db.

Net Loss Versus Time

The net loss of the undersea system (cable plus repeaters) has decreased slowly since the system was laid. The decrease is approximately cable shape, proportional to the square root of frequency. In 11 months, it has amounted to about 5 db at the top frequency. In the early months, the change was almost directly proportional to time, but later the rate slowed, as shown in Fig. 14.

The changes shown in the figure are due partly to change in cable temperature and partly to slow aging of the cable. Detailed studies of repeater crystal frequency changes have resulted in only an approximate separation into temperature effects and aging. However, it seems reasonable to assume that at the end of a 1-year cycle there would be little, if any, net change in the cable temperature averaged over the whole cable length. The greatest possibility of change would be in the shallow-water sections, and crystal measurements indicate that at the end of a year the average temperature of the shallow-water sections returned nearly to its initial value. Hence about 5 db seems chargeable to one year's aging. Extrapolation into the future, however, is uncertain. Theoretical considerations led to the idea that the rate of cable aging should decrease.

Information on long-term changes in transmission loss of earlier cables is very limited. The Havana–Key West cables were accurately measured just after laying, and again 5 years later. The change in that system from aging is very small, if indeed there is a change. That cable is generally like the transatlantic one, though it is smaller and has perhaps a significant difference in construction of the central conductor.

The foregoing applies to the net loss of the undersea part of the system only. The group-to-group net loss variation with time of the Clarenville-Oban link has been held within a much smaller range by temperature equalizer adjustments. Practices have been worked out which it is believed will hold the in-service variation with time to a fraction of a decibel.

Net Loss Versus Carrier Frequency Power Level

Single-frequency tests of carrier frequency output power versus input test power were made, up to a test power a little below the estimated overload of the highest-level repeater. An increase of 0.1 db in system net loss occurs at a power 2 to 3 db below the load limit of the transmitting amplifier. This is about 15 or 16 db above the expected rms value of the in-service system busy-hour load. The 0.1-db change is presumably due to the cumulative effect of smaller changes in several of the undersea amplifiers.

Net Loss Versus D-C Cable Current

Changes in cable current affect the repeater gain to a slight extent, the amount depending on the magnitude and phase of

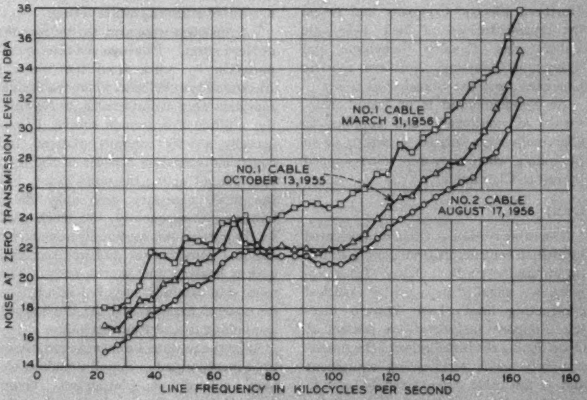


Fig. 16. Cable system noise

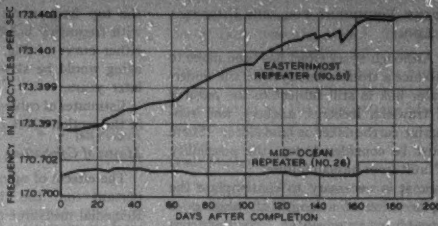


Fig. 17. Resonant frequency versus time: shore and deep-sea crystals

the feedback in the repeater as a function of frequency. Measured changes in the loss of the 2,000-mile system, for currents of 5 and 10 milliamperes less than the normal value of 225 milliamperes, are shown in Fig. 15. Under normal conditions the automatic control will hold the cable current variation within ±0.5 milliampere.

The shape of the curve in Fig. 15 is almost the same as that computed in advance from laboratory measurements on model repeaters.

SYSTEM NOISE

Values of noise on the no. 1 cable system measured in the Fall of 1955 and again in the Spring of 1956, are shown in Fig. 16. The noise increase is compatible with the decrease in undersea system net loss during this period. To prevent overload, the loss in the transmitting temperature equalizer has to be increased as the undersea loss decreases; this lowers the levels of the various parts of the undersea system by various amounts.

The noise shown in the top channel exceeds the 36-dba objective by a small amount. The excess can be recovered, if necessary, by equalization and level changes in the terminals, without recourse to compandors.

Fig. 16 shows also the noise on the no. 2 cable system shortly after completion of laying. The noise is lower than on the no. 1 cable. This is because results of experience on the no. 1 cable were utilized in better choice of cable repeater section lengths, and in better equalization while the 1956 cable was being laid.

MODULATION TESTS

Two-tone modulation tests of second- and third-order products were made, with the use of a large number of successive frequency combinations. The highest-level modulation products were at least 60 db below the 1-milliwatt test tones, at zero transmission level. This is approximately the value computed before the system was laid. Most of the modulation products were down substantially more than 60 db. Probably various causes contributed to the good performance, including the effect of misalignment in lowering repeater levels, and small propagation-time differences which minimize in-phase addition of third-order products from successive repeaters.

TELEGRAPH TRANSMISSION TESTS

At present writing, telegraph tests have been made only on the printer (telegraph order wire) channels, and without a system multichannel load. With the proposed specific telegraph level (STL) of −30 db (i.e., telegraph signal of −30 dbm per telegraph channel at 0.00-db telephone transmission level), the telegraph distortion was too low to measure reliably; it was possible to send clear messages under test conditions with a signal 36 db weaker than this.

CRYSTAL TESTS

All of the peaks of noise at the crystal resonance frequencies were easily discernible.

The crystal gain values all lay in the range from 23.6 to 27.2 db, with 60% of them lying in the range from 25 to 26 db. Crystal gain, as used here, is the difference between the system gain at a repeater crystal frequency and the average of the gains at 50 cycles above and below this frequency; the latter value is approximately the gain if the crystal were absent. No significant changes in crystal gain have occurred in 11 months of system life, and none were expected. These measurements are to be continued over the years, as an indication of electron tube aging, as explained in the companion paper on the repeater.[6]

A series of measurements of the frequency of each of the 51 repeater crystals versus time has been made. (Any of these frequency determinations can be checked on the same day within ±0.1 cycle with the special test apparatus and techniques used.) The crystals were designed to be extremely stable in frequency, so that measurements on one repeater, made at the land terminal, would not be affected by the combined effect of 50 other crystals at 100-cycle spacings. The crystal frequency varies only about −1/2 cycle per degree F increase in sea-bottom temperature. The change in frequency due to crystal aging in 11 months undersea is considered to lie in the range 0.00 cycle to −0.4 cycle.

Although the crystals were not designed as sea-bottom temperature indicators and are within a repeater housing, it seems likely that with precise techniques and after further stabilization, some uniquely accurate information on change of sea-bottom temperature will be obtainable.

In accordance with previous oceanographic knowledge of sea-bottom temperature, the frequency changes have been larger in the crystals near shore. The greatest change has been in the repeater nearest Oban. This is about 17 miles from shore and in water only about 50 fathoms deep. Its frequency change, and that of a typical deep-sea repeater, are shown in Fig. 17.

Future Considerations

SPARE EQUIPMENT FOR CABLE STATIONS

In a system as far-flung geographically as this one, and as important, much thought must be given not only to the supplies needed for routine replacement of expendable items but also to major replacements necessitated by fire or other causes. Accordingly, a schedule of spare equipment has been established, divided into two groups: "shelf" and "casualty."

Shelf spares are carried in the station itself, and include items such as dial lights, electron tubes, and dry-cell batteries, which have limited life in normal service. These are maintained in the cable station in quantity estimated as adequate for 2 years of operation.

Casualty spares embrace those major, essential frames and equipments without which the system cannot be operated. There are two subdivisions: those which are in common use in the telephone plant and which are therefore available by "cannibalization," and those which are special to the project, like the cable-current supplies and the cable-terminating bays. The common-use items are not stocked as casualty spares. Spares of the special items have been built and are stored in locations remote from the cable stations. In event of catastrophe, these can be drawn out and flown as near the point of need as possible, for onward shipment by available means.

Spare Equipment for the Undersea Link

Although every effort has been made to produce a trouble-free system, the underwater link is still subject to the hazards of trawlers, icebergs, anchors, and submarine earthquakes or landslides. There must be considered, too, the possibility of fault from human failure.

So it is necessary to contemplate the replacement of a length of cable, or a repeater, or an equalizer. For such contingency, spare cable of the various armor types has been stored on both sides of the Atlantic. Spare repeaters and equalizers are also available.

In addition, a spare called a "repair repeater" has been stocked. Need for this arises from the fact that except in very shallow water, a repair cannot be effected without the addition of cable over and above the length which was in the circuit initially. The amount of cable which must be added is a function of water depth, the condition of the sea at the time of the repair, and the amount of cable slack available in the immediate vicinity of the point in question.

When excess cable is introduced in amount sufficient to reduce significantly the system operating margins, its loss must be compensated. Hence, the repair repeater is stocked.

A repair repeater is a 2-tube device, essentially like a regular repeater, though its impedances are designed to match the cable impedances at input and output ends. However, its gain is sufficient to offset only about 5.3 miles of cable. A second type of repair repeater is under consideration, to compensate for about 15 miles of cable.

Long-Term Aging

General

In a system with some 3,200-db gross loss at its top frequency between points which are accessible for adjustment, a long-term change in loss of only 1% would have a profound effect on system performance.

For this reason, the repeater design included careful consideration of net gain change over the years.[8] The degree of control over aging is such that in a period of at least 20 years, and perhaps much longer, the estimated change in 51 repeaters might total 8-db added gain at the top frequency. The gain variation with frequency would be proportional to either curve of Fig. 15, and the rate of aging would be slower in earlier than in later years.

Estimates of cable-aging are discussed in the section "Net Loss Tests."

Means of Combatting Aging

The effects of aging would become important on the top channels first. Remedial measures to improve signal-to-noise, especially in the top channels, include: possible increase of transmission level at input of the final transmitting and load-limiting amplifier in the transmitting terminal; predistortion ahead of this amplifier; compandors; increase of direct current; and undersea re-equalization in later years. This last would be very expensive, and so it is necessary to examine the possibilities of the other measures.

The penalty for increasing the transmission level at the input of the transmitting amplifier is more peak-chopping and modulation-noise, peaks. The improvement that could be realized in this way is probably fairly small.

Predistortion is accomplished by inserting ahead of the transmitting amplifier a suitable shaping network adjusted for gain in the top part of the band and loss at lower frequencies. A complementary network (restoring network) is placed at the receiving terminal. This measure would improve signal-to-noise in the uppermost part of the band and reduce it in lower channels which have less noise. Some 3-db improvement might be thus realized in the top channel.

Compandors would give an effective signal-to-noise improvement of up to about 15 db for message telephone service, but none for services such as voice-frequency telegraph. Compandors would be applied to those channels needing them. They halve the range of talker volume, but also double the transmission variations between compressor and expandor, and thus tend to require some increases in the over-all channel net loss. The program (music) channels are already equipped with compandors which use up a part of the obtainable advantage.

If the combination of such measures should net an effective message signal-to-noise improvement of 20 db in the top channel, this would counterbalance aging of some 28 db in this channel, provided the aging were uniformly distributed along the system length. Thus considerable aging could be handled without undersea modification.

Acknowledgments

A system of the complexity of the one described obviously results from teamwork by a very large number of individuals. However, no paper on this subject could be written without acknowledgment to Dr. O. E. Buckley, J. J. Gilbert and O. B. Jacobs, now retired from Bell Telephone Laboratories. All of the early and fundamental Bell System work on repeatered submarine cable systems, and the concept of the flexible repeater, came from these sources and from their co-workers. Mr. Gilbert and Mr. Jacobs have also contributed to the present project.

References

1. Transatlantic Telephone Cable System—Planning and Over-all Performance, **E. T. Mottram, R. J. Halsey, J. W. Emling, R. G. Griffith.** Page 842 of this issue.

2. Power-Feed Equipment for the North Atlantic Link, **G. W. Meszaros, H. H. Spencer.** Page 890 of this issue.

3. Air Force Submarine Cable System, **P. T. Haury, L. M. Ilgenfritz.** *Bell Laboratories Record*, New York, N. Y., vol. 34, no. 9, Sept. 1956, pp. 321–24.

4. A Submarine Telephone Cable with Submerged Repeaters, **J. J. Gilbert.** *Bell System Technical Journal*, New York, N. Y., vol. 30, Jan. 1951, pp. 65–87.

5. Cable Design and Manufacture for the Transatlantic Submarine Cable System, **A. W. Lebert, H. B. Fischer, M. C. Biskeborn.** Page 907 of this issue.

6. Repeater Design for the North Atlantic Link, **T. F. Gleichmann, A. H. Lince, M. C. Wooley, F. J. Braga.** Page 865 of this issue.

7. Application of Compandors to Telephone Circuits, **C. W. Carter, Jr., A. C. Dickieson, D. Mitchell.** *AIEE Transactions*, vol. 65, 1946, pp. 1079–86.

8. Load-Rating Theory for Multichannel Amplifiers, **B. D. Holbrook, J. T. Dixon.** *Ibid.* (*Electrical Engineering*), vol. 59, May 1940, pp. 265–72. (Also *Bell System Technical Journal*, vol. 18, Oct. 1939, pp. 624–44.)

9. Cross Modulation in Multi-Channel Amplifiers, **W. R. Bennett.** *Bell System Technical Journal*, vol. 19, Oct. 1940, pp. 587–610.

10. Non-linear Distortion in Transmission Systems, **R. A. Brockbank, C. A. Wass.** *Proceedings*, Institution of Electrical Engineers, London, England, Mar. 1954.

11. Route Selection and Cable Laying for the Transatlantic Cable System, **J. S. Jack, W. H. Leech, H. A. Lewis.** Page 944 of this issue.

12. Electron Tubes for the Transatlantic Cable System, **J. O. McNally, G. H. Metson, E. A. Veazie, M. F. Holmes.** Page 898 of this issue.

13. Repeater Production for the North Atlantic Link, **H. A. Lamb, W. W. Heffner.** Page 877 of this issue.

Vol. 7
Pt. 1
1956

Repeater Design for the North Atlantic Link

T. F. GLEICHMANN A. H. LINCE M. C. WOOLEY F. J. BRAGA

Synopsis: Some of the considerations governing the electrical and mechanical design of flexible repeaters and their component apparatus are discussed in this paper. The discussion includes description of the feedback amplifier and the sea-pressure resisting container that surrounds it. Examples are given of some of the extraordinary measures taken to ensure continuous performance in service.

REPEATERS for use in transatlantic submarine telephone cable system had to be designed to resist the stresses of laying, and to withstand the great pressures of water encountered in the North Atlantic route. In anticipation of the need for such a long telephone system in deep water, development work was started more than 20 years ago on the design of a flexible repeater that could be incorporated in the cable and be handled as cable by conventional cable ship techniques. Successful completion, in 1950, of the design and construction of the 24-channel Key West, Fla.,–Havana, Cuba, system,[1] led to the adoption of similar repeaters designed for 36 channels for the North Atlantic link discussed in companion papers.[2,3]

Repeater transmission characteristics determine, to a large extent, the degree to which system objectives can be met. In this repeater, significant characteristics are:

1. *Noise and Modulation.* These were established by the circuit configuration and by the use of the conservative electron tube[4] developed for the Key West–Havana project.
2. *Initial Misalignment,* or mismatch of repeater gain and cable loss throughout the transmitted band of frequencies. A match within 0.05 db (decibel) was the objective. This affected both the design and the precision required in manufacture.
3. *Aging.* As electron tubes lose mutual conductance with age, repeater feedback decreases, repeater gain changes, and misalignment is affected. Decrease in feedback increases the gain at the higher frequencies so that the signal input must be reduced to prevent overloading, resulting in a signal-to-noise penalty. Gain increase is inversely proportional to the amount of feedback; in these repeaters, 33 to 34 db of feedback was the objective to keep this source of misalignment in bounds.

Because repeaters are inaccessible for maintenance, facilities are provided to enable the individual repeater performance to be checked from the shore end. This feature also permits a defective repeater to be identified in the event of transmission failure.

Repeater Unit

The repeater, for the sake of discussion, may be divided into two parts: 1. the repeater unit, which contains the electron tubes and other circuit components, and 2. the waterproof container and seals which house the repeater unit.

CIRCUIT

The circuit of the repeater unit is shown in Fig. 1. It is a 3-stage feedback amplifier of conventional design with the cathodes at a-c ground. The amplifier is connected to the cable through input and output coupling networks. Each coupling network consists of a transformer and gain-shaping elements and a power-separation inductor.

The coupling networks directly affect the insertion gain as do the two feedback networks. The design of these networks controls the insertion gain of the amplifier. The required gain (inverse of cable loss) is shown in Fig. 2. The 39-db shaping required between 20 and 164 kc is divided approximately equally among the input and output coupling networks and the feedback networks.

The interstage networks are of conventional design. The gain of the first interstage is approximately flat across the band. The second interstage has a sloping characteristic, the gain increasing with frequency. The gain shaping of these networks offsets the loss of the feedback networks so that the feedback is approximately flat across the band.

Plate and heater power is supplied to the repeater over the cable.[5] The plate voltage, approximately 52 volts, is obtained from the drop across the heater string. The d-c circuits are isolated from the container by the high-voltage blocking capacitors C_1, C_2, and C_3.

GAIN FORMULA

The circuit of Fig. 1 may be represented by a simplified circuit consisting of an input coupling network, a 3-stage amplifier, an output network, and a feedback impedance Z_β as shown in Fig. 3. From this figure it can be shown that the insertion gain of the repeater is given by[6]

$$e^{\theta}=\frac{2e^{\theta_1}e^{\theta_2}Z_c}{Z_\beta}\left(\frac{\rho_i\rho_o g_{mT}Z_1Z_2Z_\beta}{1-\rho_i\rho_o g_{mT}Z_1Z_2Z_\beta}\right) \qquad (1)$$

when $Z_\beta \ll g_{m3}Z_pZ_gZ_1Z_2 \gg (Z_o+Z_p)$ and where

$$\rho_i=\frac{Z_g}{Z_i+Z_g}$$

and

$$\rho_o=\frac{Z_p}{Z_o+Z_p}$$

are "potentiometer terms." The gain of the input network is defined as $e^{\theta_1}=V/E_i$; where V is the open-circuit voltage of the input network with E_i as the source, and the gain of the output network is defined as $e^{\theta_2}=i/I_1$. This expression may be put in familiar form by recognizing that $\rho_i\rho_o g_{mT}Z_1Z_2Z_\beta$ is $\mu\beta$, the feedback around the loop. Hence

$$e^{\theta}=\frac{2e^{\theta_1}e^{\theta_2}Z_c}{Z_\beta}\left(\frac{\mu\beta}{1-\mu\beta}\right) \qquad (2)$$

Equation 2 shows that the insertion gain of the repeater is the product of five factors, namely:

1. e^{θ_1}, the gain of the input network.
2. e^{θ_2}, the gain of the output network.
3. Z_c, the cable impedance.
4. Z_β, the feedback impedance.
5. $(\mu\beta/1-\mu\beta)$, the $\mu\beta$ effect term.

It should be noted that a number of simplifying assumptions have been made. For example, the effect of grid-plate capacitance has been neglected. In addition, the β circuit has been assumed to be a 2-terminal impedance whereas it is actually a 4-terminal network. However, in the passband and over a large part of the outband of the repeater, these simplifications give a very good approximation to the true gain of the repeater.

In the passband, $(\mu\beta/1-\mu\beta)$ is very nearly unity so that the gain controlling factors are e^{θ_1}, e^{θ_2}, and $1/Z_\beta$, assuming that Z_c is fixed.

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957.

T. F. GLEICHMANN, A. H. LINCE, M. C. WOOLEY, and F. J. BRAGA are with the Bell Telephone Laboratories, Murray Hill, N. J.

Material for the section, "Power By-pass Gas Tube," was contributed by M. A. Townsend.

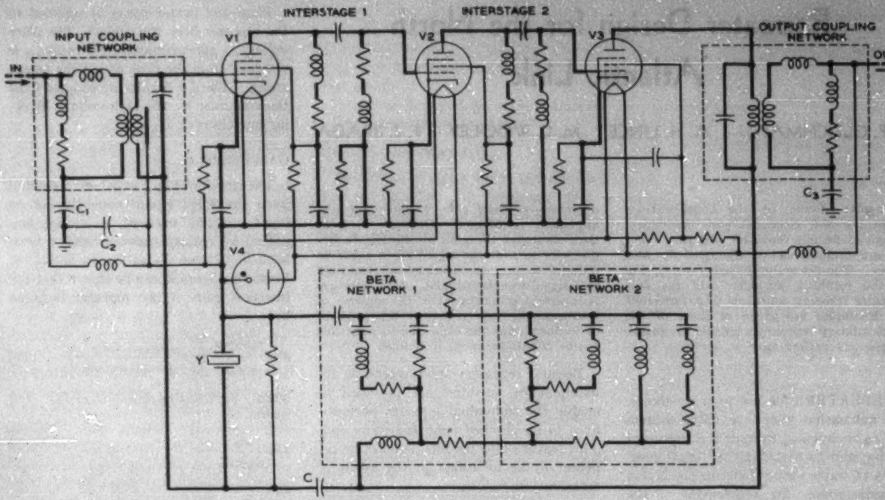


Fig. 1. Repeater schematic

COUPLING NETWORKS

The input and output networks are essentially identical. The networks are of unterminated design and therefore do not present a good termination to the cable at all frequencies, which results in some ripple in the system transmission characteristic at the lower edge of the band, and makes the repeater insertion gain sensitive to variations in the cable impedance. However, this arrangement has the advantage of maximum signal-to-noise performance, highest gain, and most effective shaping with a minimum of elements. A minimum of elements is important in view of the space restrictions imposed by the flexible repeater structure. The sensitivity of the gain to variation in impedance is minimized by close manufacturing control of the cable and networks.

The schematic of a coupling network is shown in Fig. 4, and the equivalent circuit in Fig. 5. Capacitor C_1 and inductor L_4 are part of the power-separation circuit. The effect of L_4 in the transmission band is negligible, and it has been omitted from the equivalent circuit. However, C_1 is in the direct transmission path and has a small effect at the lower edge of the band so that it becomes a design parameter. The combination R_2, L_2 controls the low-frequency gain shaping of the network. Inductor L_3 builds out the leakage inductance of the transformer and together with capacitor C_3 controls the shaping at the top end of the band. These elements are adjusted during manufacture of the networks to provide the desired shaping.

The equivalent circuit is an approximation to the true transformer circuit. By standard network analysis techniques the ratio V/E_1, the gain of the network, can be obtained. The agreement between measurements and computation is sufficiently close, several hundredths of a decibel, to ensure that the representation is good.

Each coupling network is designed to provide approximately one third of the total shaping required, or 13 decibels. While these networks are outside the feedback path, the impedances which they present to the amplifier are important factors in the feedback design. It can be seen from Fig. 3 that, at the amplifier input, the proportion of the feedback voltage which will be effective in producing feedback around the loop is dependent upon the potentiometer division between the grid-cathode impedance of the first tube and the impedance looking back into the coupling network. The greater the gain shaping of the network, the greater the potentiometer loss. The maximum gain which can be obtained from the coupling network is limited by the capacitance across the circuit. This capacitance cannot be reduced without increasing the potentiometer loss, and seriously limiting feedback. In this design, an acceptable compromise is made when the ratio of network capacitance to grid-cathode capacitance has been fixed at 1.2, as suggested by Bode.[8] The gain through the input network and the deviation from one-third cable shape is shown in Fig. 6. A typical potentiometer term is shown in Fig. 7.

Similar considerations apply to the output network with the further restriction that the impedance presented to the output tube should be about 40,000 ohms at the top edge of the band for optimum modulation performance.

The coupling networks have a temperature characteristic which must to taken into account in the insertion gain of the repeater. The characteristic is due to variations in the resistance of C_1 and R_2 with temperature. This amounts to 0.005 db per F (degree Fahrenheit) at 20 kc, decreasing with frequency and becoming negligible above 80 kc.

BETA CIRCUIT

The beta or feedback network is designed to complement the combined characteristics of the input and output coupling networks and mop-up residual effects, such as those due to $\mu\beta$ effect and coupling-network temperature coefficients. The network also provides the d-c path for the output-tube plate current.

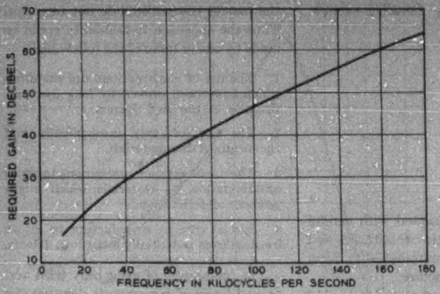


Fig. 2. Required insertion gain

The configuration of the beta circuit is shown in Fig. 8. In the passband it is a 2-terminal network whose impedance varies from about 300 ohms at 20 kc to 70 ohms at 164 kc. It consists of essentially two parts. The elements to the left of the dotted line provide the major portion of the shaping. With these, the repeater is within ±0.7 db of the required gain. The series resonant circuits to the right of the dotted line reduce this to the ±0.05-db set as the objective.

The mopping-up elements are connected to the main portion of the beta circuit through a resistance potentiometer R_1, R_2, and R_3. This scales the elements of the resonant circuits to values which would meet mounting space and component restrictions.

BUILT-IN TESTING FEATURES

The crystal *Y* and capacitor *C*, Fig. 1, in the feedback path provide the means for checking the repeater from the shore station. The crystal is a sharply tuned series resonant shunt on the feedback path, which reduces the feedback at the resonant frequency and thus produces a narrow peak in the insertion gain characteristic of the repeater. The feedback reduction, and hence the peak gain, is controlled by the potentiometer divider formed by the reactance of the capacitor and the series resonant resistance of the crystal. The crystal and capacitor are chosen so that substantially all the feedback is removed from the repeater. With no feedback, the peak gain is proportional to the mutual conductance of the three tubes.

At frequencies well off resonance, the impedance of the crystal is high so that no reduction in feedback results. Periodic measurements of gain, at the resonant frequency relative to measurements made at a frequency off resonance, will show any changes in the tubes. The crystal frequency is different for each repeater so that by measuring the gain from the shore stations at the various crystal frequencies it is possible to monitor the performance of the individual repeaters.

The increase in gain at the peak is approximately 25 db. The crystal frequencies, spaced at 100-cycle intervals, are placed above the normal transmitted band between 167 and 173.4 kc.

Thermal noise, always present at the input to the repeater, is also amplified over the narrow band of frequencies corresponding to the peak gain in each repeater, so that at the receiving end of the line there are a series of noise peaks, one for each repeater. Should a repeater fail, the noise peaks of all repeaters between the faulty repeater and the receiving end will be present and those from repeaters ahead will be missing. By determining which peaks are missing, the location of the failed repeater can be determined. It is obvious that to locate a faulty repeater the power circuit must be intact. To guard against power interruption owing to an open electron-tube heater, a gas tube *V*4, Fig. 1, is connected across the heater string as a by-pass.

LOOP FEEDBACK

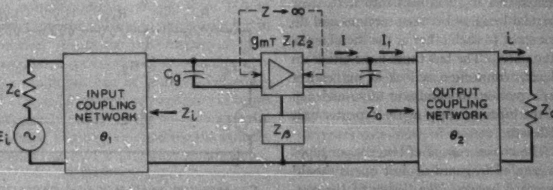
The design of the feedback loop follows conventional practice. The restrictions that limit the amount of feedback that can be obtained in the transmitted band are well known.[6] Broadly speaking, the figure of merit of the electron tubes and the incidental circuit capacitances determine the asymptotic cutoff which limits the amount of feedback that can be obtained in the band. With the flexible repeater circuit, capacitances are rather large because of the severe space restrictions and physical length of the structure. Transit time of 1.8 degrees per megacycle per tube and a like amount for the physical length of the feedback loop reduced the available feedback by 2 db.

Margins of 10 db at phase crossover and 30 degrees at gain crossover were set as design objectives. While these may seem to be ultraconservative in view of the tight controls placed on components and the mechanical assembly, it should be borne in mind that the repeaters are inaccessible and repairs would be costly.

Modulation and tube-aging considerations require a minimum feedback of 33 to 34 db. With the restrictions noted previously and the effect of the potentiometer terms on the available feedback, the top edge of the band is limited to about 165 kc with the desired feedback.

MECHANICAL DESIGN

To provide a flexible structure the repeater unit is assembled in a number of longitudinal sections mechanically coupled by helical springs and electrically interconnected by means of bus tapes. The assembly is composed of 17 sections. Figs. 9 and 10 show the repeater make-up and an assembled unit.



V = OPEN CIRCUIT VOLTAGE OF INPUT COUPLING NETWORK WITH E_i AS THE SOURCE

θ_1 = GAIN OF INPUT COUPLING NETWORK DEFINED AS $e^{\theta_1} = V/E_i$

θ_2 = GAIN OF OUTPUT COUPLING NETWORK DEFINED AS $e^{\theta_2} = i/I$

Z_1Z_2 = INTERSTAGE IMPEDANCES

g_{mT} = PRODUCT OF g_m OF THREE AMPLIFIER TUBES

Fig. 3. Simplified amplifier circuit

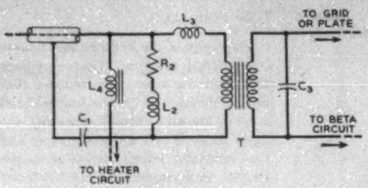


Fig. 4. Coupling network

The sections consist of the circuit component, or components, mounted in machined plastic forms and enclosed in a plastic container which is enclosed in a housing of the same material. The material used is methyl methacrylate which was chosen for its physical and chemical stability and good machinability. The sections contain circuit components, grouped functionally, such as input coupling network, interstage, electron tube, or high-voltage blocking capacitor. In the case of the feedback network it was necessary to mount the network in two sections because of the large numbers of components involved. A typical network, container, and housing are shown in Fig. 11.

The bus tapes are placed in grooves milled in the outer surfaces of the section containers. Wiring spaces are machined into the ends of the containers for connecting network leads to the bus tapes. The housing is placed over the container and the busses and is closed by a plastic coupler plate which also forms part of the intersection couplers. The coupler plates are fastened to the housing with plastic pins.

Between sections, the bus tapes are looped toward the longitudinal axis of the repeater unit. The dimensions of the loop are rigidly controlled so that, as the unit is flexed during bending of the repeater, the loops always return to their original location between sections and do not short to each other or the metal outer container. The bus tapes have either an electric connection or lock at one end of each section to eliminate any tendency of the tapes to creep as the repeater unit is flexed.

The busses consist of two copper tapes in parallel to guard against opens should one tape break. The design of the connections to the busses is such that once the section is closed there can be no disturbance of the tapes or network leads in the vicinity of the electric connections. The bus-type wiring plan was chosen as the best arrangement for the long structure in keeping with the stringent transmission requirements. Electrically adjacent but physically remote components can thus be interconnected with careful control of the parasitic capacitances and couplings to insure reproducibility from unit to unit in manufacture.

Components

The development of passive components for use in the flexible repeater presented a number of unusual problems, the most important being: 1. the extreme reliability, 2. the high degree of stability, 3. the limitations on size and shape, and 4. an environment of constant low temperature.

The repeaters for the transaltantic system contain a total of approximately 6,000 resistors, capacitors, inductors, and transformers. If we are to be 90% (per cent) certain of attaining the objective of 20 years, service without failure of any of these components, the effective average annual failure rate for the components must not be more than one in a million. To assure this degree of reliability by actual tests would require more than 400 years of testing on 6,000 components. Obviously, some other approach to insure reliability is required. The most obvious avenue, that of providing a large factor of safety, was not open because of space limitations.

Fortunately, with only one exception, the passive components do not wear out. Thus the approach to reliability could be made by one or more of the following:

1. The use of constructions and materials which have been proved by long use, particularly in the Bell System.
2. The use of only mechanically and chemically stable materials.
3. The use of extreme precautions to avoid contamination by materials which might promote deterioration.
4. Special care in manufacture to insure freedom from potentially hazardous defects.

The philosophy of using only tried and proved types of components dictated the use of wire-wound resistors, impregnated-paper and silvered-mica capacitors, and permalloy cores for inductors and transformers. While newer and, in some ways, superior materials are known, none of these possessed the necessary long record of trouble-free performance. In some cases, particularly in resistors, this approach resulted in more difficult design problems and also in physically larger components. While the ambient conditions in the repeater, i.e., low temperatures and extreme dryness, are ideal from the standpoint of minimizing corrosion or other harmful effects of a chemical nature, the materials used in the fabrication of components were nevertheless limited to those which are inherently stable and nonreactive. In addition, raw materials were carefully protected from contamination from the time of their manufacture until they were used, or, wherever possible, they were cleaned and tested for freedom from contaminants just prior to use. Unusually detailed specifications were prepared for all materials.

The effort to achieve extreme reliability also influenced or dictated a number of

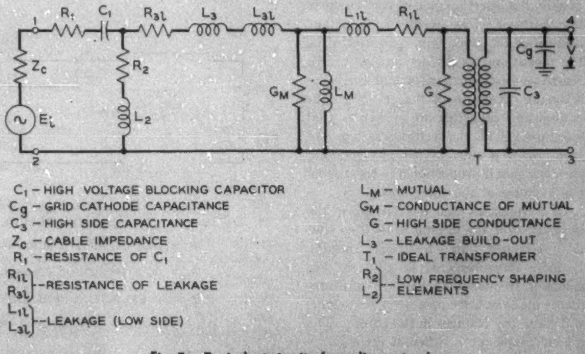

1
R1
C1
R3L
L3
L3L
L1L
R1L
4
Zc
R2
GM
LM
G
Cg
V
C3
Ei
L2
2
T
3
C1 – HIGH VOLTAGE BLOCKING CAPACITOR
Cg – GRID CATHODE CAPACITANCE
C3 – HIGH SIDE CAPACITANCE
Zc – CABLE IMPEDANCE
R1 – RESISTANCE OF C1
R1L, R3L – RESISTANCE OF LEAKAGE
L1L, L3L – LEAKAGE (LOW SIDE)
LM – MUTUAL
GM – CONDUCTANCE OF MUTUAL
G – HIGH SIDE CONDUCTANCE
L3 – LEAKAGE BUILD-OUT
T1 – IDEAL TRANSFORMER
R2, L2 – LOW FREQUENCY SHAPING ELEMENTS

Fig. 5. Equivalent circuit of coupling network

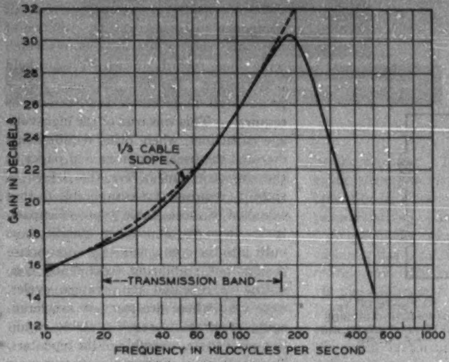


Fig. 6. Gain of input coupling network

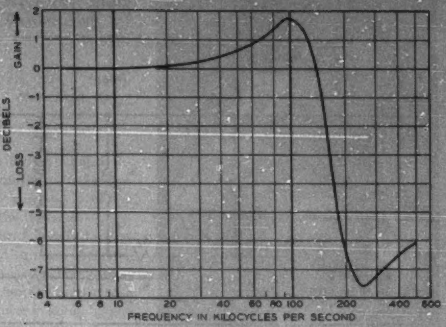


Fig. 7. Input potentiometer term

design factors such as the minimum wire diameters used in wound apparatus, the use of as few electrical joints as possible, and the use of relatively simple structures. These limitations resulted in the use of unencased components in most instances. Wherever possible, the ends of windings were used as terminal leads to avoid unnecessary soldered connections. This injected the additional hazard of lead breakage owing to handling during manufacture and inspection. This hazard was minimized in most instances by providing the windings with extra turns which were removed just before the component was assembled in the network. Thus, the lead wires in the final assembly had never been subjected to severe stress. Where this technique was impracticable, special fixtures and handling procedures were used to prevent undue flexing or stressing of lead wires.

As mentioned previously, there was one type of passive component in which life is a function of time and severity of operating conditions. These are the capacitors, especially those subject to high voltages. Because of this and the fact that the physical and electrical requirements dictated the use of relatively high dielectric stress in these capacitors, a program of study covering a wide range of dielectric materials was undertaken about 1940. This study showed that none of the usual solid or semisolid materials used to impregnate paper capacitors were suitable for continuous use at sea-bottom temperatures. Typical results of this program are shown in Figs. 12 and 13. These curves show the performance of capacitors operating at approximately 1.8 times normal dielectric stress at both sea-bottom and room temperatures. It is evident that even semisolid impregnants are inferior to liquids at the lower temperature. The need for the maximum capacitance in a given space restricted the field still further, so that the final choice was a design using castor-oil-impregnated kraft paper as the dielectric.

It is well established that the life of impregnated-paper capacitors is inversely proportional to the fourth to sixth power of the voltage stress or

$$\frac{L_1}{L_2}=\left(\frac{V_2}{V_1}\right)^p$$

where p ranges from 4 to 6. This fact permits the accumulation of a large amount of life information in a relatively short time. To insure that the capacitor design selected would provide the degree of reliability required, a number of capacitors were constructed and placed on test at voltage stresses ranging from $1^1/_2$ to $2^1/_4$ times the maximum stress expected in service. From the performance of these samples, a prediction of that under service conditions can be made as follows.

The total equivalent exposure, in terms of capacitor years at the maximum service voltage, can be computed for the samples under test by the following summation

$$T=N_1T_1\left(\frac{V_1}{V_s}\right)^p+N_2T_2\left(\frac{V_2}{V_s}\right)^p+\dots+ N_rT_r\left(\frac{V_r}{V_s}\right)^p \quad (3)$$

where N_1, N_2, … N_r are the number of samples on test at voltage stresses V_1, V_2, V_r; and T_1, T_2, … T_r are the total times of the individual tests, and V_s is the maximum voltage stress under service conditions.

If, as has been the case in the tests described, there has been only one failure in the total exposure T, we can estimate, from probability equations, the limits or bounds within which the first failure will occur in a system involving a given number of capacitors, operating at a voltage stress V_s. These equations are probability of no failures in exposure

$$T=e^{-(T/L)} \quad (4)$$

probability of more than one failure in exposure time

$$T=1-\left(1+\frac{T}{L}\right)e^{-T/L} \quad (5)$$

where T is obtained from equation 3, and L is the total exposure in the same units as T for the service conditions. The solutions of equations 4 and 5 for L, using any desired probability, give the maximum and minimum exposures in capacitor-years, within which the first failure may be expected to occur under service conditions.

However, since the voltage on the capacitors varies from repeater to repeater, it is necessary to determine the equivalent exposure of the system, in terms of capacitor-years per year of operation at the maximum service voltage, to estimate the time to the first failure in the system. This is obtained from equation 3, for one half of one cable, by substituting the supply voltage at each repeater for V_1, V_2, etc., the maximum service voltage for V_s, and the number of capacitors per repeater for N. The total exposure for a 2-cable system is then four times this figure. With the data which have been accumulated and the number of capacitors and voltages of the transatlantic system, it is

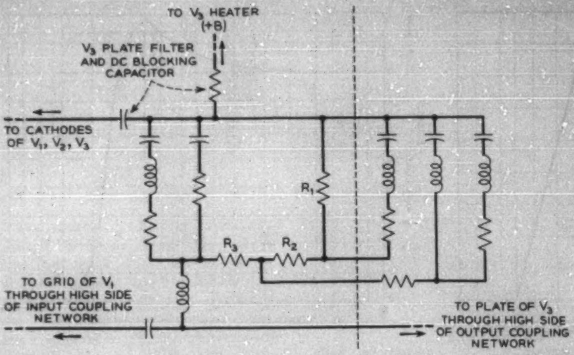

TO V_3 HEATER (+B)
V_3 PLATE FILTER AND DC BLOCKING CAPACITOR
TO CATHODES OF V_1, V_2, V_3
R_1
R_3
R_2
TO GRID OF V_1 THROUGH HIGH SIDE OF INPUT COUPLING NETWORK
TO PLATE OF V_3 THROUGH HIGH SIDE OF OUTPUT COUPLING NETWORK

Fig. 8. Beta network

estimated, with a probability of being correct nine times in ten, that the first wear-out failure of a capacitor in the transatlantic system will not occur in less than 16 years nor more than 600 years.

There is, of course, the possibility of a catastrophic or early failure due to mechanical or other defects not associated with normal deterioration of the dielectric. Such potential failures are not always detected by the commonly used short-time over-voltage test. Thus, for submarine cable repeaters, all capacitors subjected to d-c potentials in service are subjected to at least $1^1/_2$ times the maximum operating voltage for a period of 4 to 6 months before they are used in repeaters. Experience indicates that this is adequate to detect potential early failures. The results of this type of testing on submarine cable capacitors is an indication of the care used in selecting materials and manufacturing the capacitors. Only one failure has occurred in more than 3,000 capacitor-years of testing.

An important aspect of the control of quality of components is the control of the raw materials used in their manufacture. For the transatlantic project, this was accomplished by rigid specifications, thorough inspection and testing, supplemented in some cases by a process of selection.

This can be illustrated by the procedure used for selecting the paper which is used as the dielectric in capacitors. The Western Electric Company, Inc., normally inspects many lots of capacitor paper during each year. Those lots which were outstanding in their ability to stand up under a highly accelerated voltage test were selected from this regular inspection process. These selected lots were then subjected to a somewhat less highly accelerated life test. Paper which met the performance requirements of this test was slit into the proper widths for use in capacitors. Sample capacitors were then prepared with this paper and so selected that they represented a uniform sampling of the lot at the rate of one sample for approximately each 3 pounds of paper. These samples were impregnated with the same lot of oil to be used in the final product. Satisfactory completion of accelerated life and other tests on these samples constituted final qualification of the paper for production of capacitors. Relatively few raw materials were adaptable to such tests or required such detailed and exhaustive inspection as capacitor paper. But the attitude in all cases was that the material be qualified not only as to its primary constituents or characteristics, but also as to its uniformity and freedom from unwanted properties.

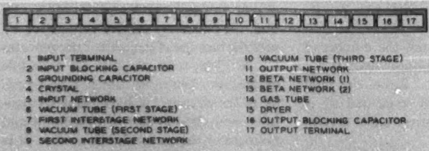

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
1 INPUT TERMINAL
2 INPUT BLOCKING CAPACITOR
3 GROUNDING CAPACITOR
4 CRYSTAL
5 INPUT NETWORK
6 VACUUM TUBE (FIRST STAGE)
7 FIRST INTERSTAGE NETWORK
8 VACUUM TUBE (SECOND STAGE)
9 SECOND INTERSTAGE NETWORK
10 VACUUM TUBE (THIRD STAGE)
11 OUTPUT NETWORK
12 BETA NETWORK (1)
13 BETA NETWORK (2)
14 GAS TUBE
15 DRYER
16 OUTPUT BLOCKING CAPACITOR
17 OUTPUT TERMINAL

Fig. 9. Repeater make-up

To a considerable extent, stability of components is assured by the practice of using only those types of structures which have long records of satisfactory field performance. However, in some cases, a product far more stable than usual was required. This was true of the high-voltage capacitors, where other requirements dictated the use of impregnated paper as the dielectric, but where the degree of stability required was comparable to that expected of more stable types of capacitors. In so far as possible, stability was built into the components by appropriate design, but stabilizing treatments, consisting of repeated temperature cycles, were used where necessary to accelerate aging processes to reach a stable condition prior to the assembly of the repeaters. Temperature cycling or observation over periods up to 6 months were used also to determine that the components' characteristics were stable. Exceptional inspection procedures which were followed to insure reliability and stability are described in detail in a companion paper.[7]

As mentioned earlier, the design and construction of components was simplified by omitting housings or containers, except for oil-impregnated paper capacitors. Adequate mountings for the components were obtained in several ways. Mica capacitors were cemented to small bases of methyl methacrylate which were in turn cemented in suitable recesses in network structures. Inductors and transformers were cemented directly into recesses in the network housings. Fig. 14 illustrates some of these structures and their mounting arrangements. On the bottom is a molybdenum permalloy dust-core coil in which a mounting ring of methyl methacrylate, provided with radial fins, is secured around the core by tape and the wire of the winding. Such inductors were mounted by cementing the projecting fins into slots arranged around a recess in the network housing. On top is an inductor which, for electrical reasons, required a core of greater cross section than could be accommodated in the network when made by the usual toroidal construction. In this case, the effective cross section of two cores was obtained by cementing the cores in a figure-eight position and applying the winding so that it threads the hole in both cores. With these constructions, the cement used to secure the inductors does not come into contact with the wire of the winding, which is thereby not subject to strains produced by curing of the cement.

For economy of space and also to reduce the number of soldered connections, many of the components' structures con-

Fig. 10 (left). Over-all view of the repeater unit

Fig. 11 (right). Network section

tain two or more elements. Inductors and resistors were combined by winding inductors with resistance wire. Separate adjustment of inductance and resistance were obtained by adjusting turns for inductance and the length of wire in a small noninductive winding for resistance. The inductor on the bottom, in Fig. 14, illustrates one type in which the noninductive part of the winding is placed on one of the separating fins. In some cases, capacitors and resistors were also combined. Fig. 15 illustrates two of these. The capacitor at the lower right contains three capacitances and a single resistance in the same container. This requires that the resistor parts be capable of withstanding the capacitor drying and impregnation process, and also that the resistor contain nothing which would be harmful to the capacitor. The capacitor on the left, in this figure, is housed in a ceramic container on which is wound a resistor. The capacitor at the top is a high-voltage type which, aside from electron tubes, represents the largest single component used in the repeater. In this capacitor, the tape terminals which contact the electrodes are brought out through the ceramic cover and are made long enough to reach an appropriate point, so as to avoid additional soldered connections. Such special designs introduced many problems in the manufacture of the components. However, the improved performance of the repeater and the increase in the inherent reliability of the over-all system fully justified the greater effort which was required for the production of such specialized apparatus.

Power By-pass Gas Tube

The fault-locating means, referred to previously, requires that the power circuit through the cable be continuous. To protect against an open circuit in the repeater, such as a heater failure, an additional device is required to by-pass the line current. This by-pass must be a high resistance under normal operating conditions since any current taken by this device muse be supplied through preceding repeaters. If an open circuit occurs, the by-pass must carry the full cable current. At full current, the voltage drop should be small to avoid excessive localized power dissipation in the repeater. The device should recover when power is removed, so that false operation by a transient condition will not permanently by-pass the repeater.

A gas diode, using an ionically heated cathode, has been used to meet these requirements. By making the breakdown voltage safely greater than the drop across the heater string, no power is taken by the tube under normal repeater operation. In the event of an open circuit in the repeater, the voltage across the tube rises and breakdown occurs. Full cable current is then passed through the gas discharge. Removal of power from the cable allows the tube to deionize and recover in the event of false triggering by transients. The cathode is a coil of tungsten wire, coated with a mixture of barium and strontium oxide. A cold-cathode glow discharge forms when the tube is first broken down. This discharge has a sustaining voltage of the order of 70 volts. The glow discharge initially covers the entire cathode area. Local heating occurs and some parts of the oxide coating begin to emit electrons thermionically. This local emission causes increased current density and further increases the local heating. The discharge thus concentrates to a thermionic arc covering only a portion of the coil. The sustaining voltage is then of the order of 10 volts.

Mechanically, the tube was designed to minimize the possibility of a short circuit resulting from structural failure of tube parts. Fig. 16 shows the construction of the tube. The glass envelope and stem structure, which had previously been developed for the hot-cathode repeater tubes, were used as a starting point for the design. The anode is a circular disk of nickel attached to two of the stem lead wires. To provide shock resistance, the supporting stem leads are crossed and welded in the center. To protect against weld failure, a nickel sleeve is used at each end of the cathode coil. It is crimped to hold the coil mechanically in place and then welded at the end for electric connection. At the end of the coil as well as in all other places where it is possible, a mechanical wrap is made in addition to spot welding. An additional precaution is taken by inserting an insulated molybdenum support rod through the center of the cathode coil. The filling gas is argon at a pressure of 10 millimeters of mercury. To provide initial ionization, one microgram of radium, in the form of radium bromide, was placed on the inside of the tube envelope. All materials were procured in batches of sufficient size to make the entire lot of tubes and were carefully tested before being approved for use. The tubes were fabricated in small groups and a complete history was kept of the processing of each lot.

For detailed study of tube performance, a number of electrical tests were made. These involved measurements of breakdown voltage, operating voltage as a glow discharge at low current, current required to cause the transition to a thermionic arc, the time required at the cable current to cause transition to the low-voltage arc, and the sustaining voltage at the full cable current.

All tubes were aged by operating at 250 ma (milliamperes) on a schedule which included a sequence of short on-off periods, 2 minutes on, 2 minutes off, followed by periods of continuous operation. A total of 150 starts and 300 hr (hours) of continuous operation were used. Following this aging schedule, the tubes were allowed to stabilize for a few days and then subjected to a 2-hr thermal treatment or pulse at 125 C (degrees centigrade). It was required that no more than a few volts change in breakdown voltage occur during this thermal pulse before a tube was considered as a candidate for use in repeaters.

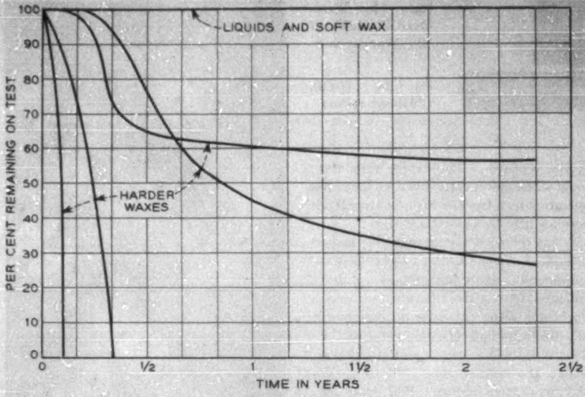


Fig. 12. Accelerated life tests on paper capacitors with various impregnants at room temperature 60 to 80 F

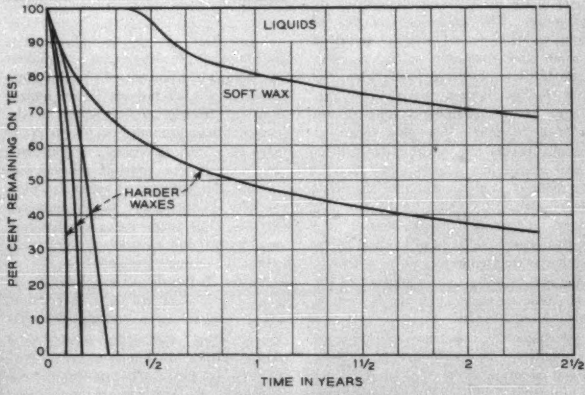


Fig. 13. Accelerated life tests on paper capacitors with various impregnants at 40 F

After aging and selection as candidates for repeaters, tubes were stored in a light-tight can at 0 C. Measurements were made to assure stability of breakdown voltage and breakdown time.

The quality of each group of 12 tubes was further checked by continuous- and on-off cycling life tests. The fact, that none of these tubes has failed on the cycling tests at less than 3,500 hr and 1,500 starts and no tube on continuous operation has failed at less than 4,200 hr, gives assurance that system tubes will start once and operate for the few hours necessary to locate a defective repeater. Long-term shelf tests of representative samples at 70 C and at 0 C give assurance of satisfactory behavior in the system.

Container and Seals

The design of the flexible enclosure for the flexible repeater unit is basically the same as it emerged from its development stages in the 1930's. It is virtually identical to the structure of the repeaters manufactured by the Bell Telephone Laboratories for the cables laid in 1950 between Key West and Havana.[1]

The functions of the enclosure are: 1. to protect the repeater unit from the effects of water at great pressure at the ocean bottom, 2. to provide means of connecting the repeater to the cable before laying, and 3. to be slender and flexible enough to behave like cable during laying. How these functions are met in the design may be more readily understood by reference to Fig. 17.

The repeater unit, described earlier, is surrounded by a 2-layer carcass of steel rings, end to end. The rings are surrounded in turn by a copper tube $1^3/_4$ inches in diameter and having a 1/32-inch wall.

When a repeater is bent during laying by passing onto the cable-ship drum, the steel rings separate at the outer periphery of the bend and the copper tube stretches beyond its elastic limit. As the repeater leaves the drum under tension, the rings separate and the copper stretches on the opposite side, leaving the repeater in a slightly elongated state. At the ocean bottom, hydraulic pressure restores the repeater to its original condition with rings abutted and the copper tube reformed.

The system of seals in each end of the tube consists of: 1. a glass-to-Kovar seal adjacent to the repeater unit, 2. a rubber-to-brass seal seaward from the glass seal, and 3. a core-tube and core-sleeve seal seaward from the rubber seal.

The glass seal, although capable of withstanding sea-bottom pressures, is primarily a water-vapor barrier and a lead through for electric connection to the repeater circuit. In service, it is normally protected from exposure to sea pressures by the rubber seal.

The rubber seal, capable of withstanding sea-bottom pressures, is indeed exposed to these pressures for the life of the repeater, but is not exposed to sea water. It is likewise a lead through for electric connection from the cable to the glass seal.

The core-sleeve seal is an elastic barrier between sea water on the outside and a fluid on the inside. This fluid, polyisobutylene, is a viscous honey-like substance, chemically inert, electrically a good insulator, and a moderately good water-vapor barrier. It fills the long thin annular space outside the cable core and inside a copper-core tube, and thus becomes the medium of transmitting to the rubber seal the sea pressure exerted on the core sleeve. It can be seen that the core-sleeve seal has nominally no pressure-resisting function and no electrical function.

The same fluid is also used to fill the space between the glass and rubber seals. Voids at any point in the system of seals

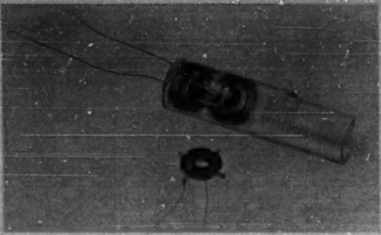


Fig. 14. Mounting for molybdenum permalloy dust-core coil

are potential hazards to long trouble-free life. Empty pockets, for instance, lying between the central conductor and the outer conductor, or container, are capable of becoming electrically conducting paths, if filled with water vapor. As pointed out in companion papers,[4,5] the voltage between the repeater and cable central and outer conductors is about 2,000 volts at the ends of the transatlantic system.

The filling of the seal interspace with a liquid would defeat one function of the rubber seal if special features were not provided in the rubber-seal design. Very slight displacement of the rubber seal toward the glass seal because of sea pressure, or resulting from reduction in volume because of falling temperature, would otherwise build up pressure in the liquid and on the glass seal. This is avoided by providing a kind of resilience in the interspace chamber. Three small brass bellows, partly compressed, occupy fixed cavities in the chamber. They can compress readily and maintain essentially constant conditions, independent of external pressures and temperatures.

The entire repeater assembly enclosed in copper is approximately 23 feet long. Tails of cable at each end make the total length about 80 feet before splicing. The central conductor of each cable tail is joined to the rubber-seal central conductor, with the insulation molded in place in generally the same manner as in cable-to-cable junctions elsewhere in the system. The outer-conductor copper tapes of the cable tails are electrically connected to the copper-core tubes.

The copper region is coated with asphalt varnish and gutta-percha tape to minimize corrosion. Over this coating, bandagelike layers of glass-fabric tape are built up to produce an outer contour tapering from cable diameter at one end up to repeater diameter, and back down to cable diameter at the opposite end. The tape covering is saturated with asphalt varnish. This tape is primarily a bedding for the armor wires that are laid on the outside of both cable tails and repeater to make the repeater cablelike in its tensile properties and capable of being spliced to cable.

In the region of the repeater proper where the diameter is double that of cable, extra armor wires are added to produce a layer without spaces. Also, to avoid subjecting the repeater to the torque characteristically present in cable under the tensions of laying, a second layer of armor wires of opposite lay is added over the first layer. This process is so closely related to the armoring of cable core in a cable factory that it is performed there.

MATERIALS

Following the same design philosophy applied to the repeater components, the materials of construction of the repeater container and seals were chosen for maximum life, compatibility with each other, and for best adaptability to the design intent. Specifications, particularly adapted to this use were set up for all of the some 50 different metals and nonmetals used in the enclosure design. In general, the methods established for proving the integrity of the materials are more elaborate than usual commercial practice. In most instances, such as that of copper container tubes, the extraordinary inspection for defects and weaknesses with its resulting rejection rate resulted in high cost for the usable material.

Testing

A substantial part of the development work on the repeater enclosure was concerned with devising tests that give real assurance of soundness and stability. It is beyond the scope of this paper to discuss how each part is tested before and after it is assembled, but certain outstanding tests deserve mention.

STEEL-RING TESTS

Each of the inner steel rings, before installation, is required to pass a magnetic-particle test to find evidence of hidden metallurgical faults. Each ring is later a participant in a group test under hydraulic pressure simulating the crushing effect of ocean-bottom service, but exceeding the working pressures. The magnetic-particle test is repeated.

HELIUM-LEAK TESTS

Both glass- and rubber-seal assemblies, before being installed in repeaters, are required to undergo individual tests

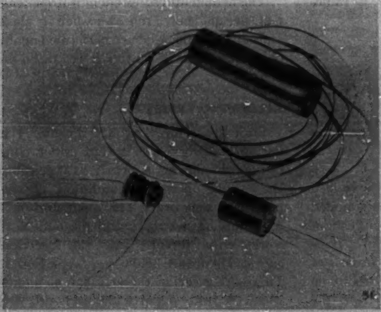


Fig. 15 (left). Capacitor and resistor - capacitor combinations

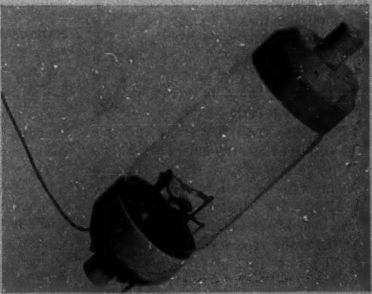


Fig. 16 (right). Power by-pass gas tube

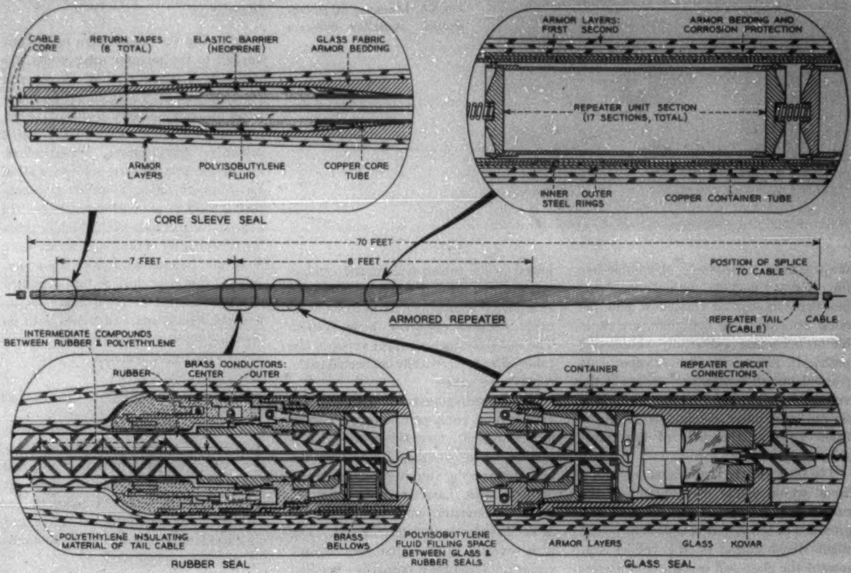

CABLE CORE
RETURN TAPES (8 TOTAL)
ELASTIC BARRIER (NEOPRENE)
GLASS FABRIC ARMOR BEDDING
ARMOR LAYERS
POLYISOBUTYLENE FLUID
COPPER CORE TUBE
CORE SLEEVE SEAL
ARMOR LAYERS: FIRST SECOND
ARMOR BEDDING AND CORROSION PROTECTION
REPEATER UNIT SECTION (17 SECTIONS, TOTAL)
INNER OUTER STEEL RINGS
COPPER CONTAINER TUBE
70 FEET
7 FEET
8 FEET
POSITION OF SPLICE TO CABLE
ARMORED REPEATER
REPEATER TAIL (CABLE)
CABLE
INTERMEDIATE COMPOUNDS BETWEEN RUBBER & POLYETHYLENE
RUBBER
BRASS CONDUCTORS: CENTER OUTER
POLYETHYLENE INSULATING MATERIAL OF TAIL CABLE
BRASS BELLOWS
RUBBER SEAL
POLYISOBUTYLENE FLUID FILLING SPACE BETWEEN GLASS & RUBBER SEALS
CONTAINER
REPEATER CIRCUIT CONNECTIONS
ARMOR LAYERS
GLASS
KOVAR
GLASS SEAL

Fig. 17. Details of container and seals

under high-pressure helium gas. Helium is used not only because its small molecules can pass through smaller leaks than can water molecules, but also because of the excellent mass spectrometer type of leak detectors commercially available for this technique. While helium is applied at high pressure to the outer wall of the seal, the inner wall is maintained under vacuum in a chamber joined with the leak detector. The passage of helium through a faulty seal at the rate of 10^{-9} milliliters per second can be detected. Stated differently, this is one milliliter of helium in 30 years. The relation of water-leak rate to helium-leak rate is dependent on the physical nature of the leak, but if they were assumed to be equal rates, the amount of water which might enter a tested repeater in 20 years would be 0.66 gram. A desiccant within the repeater cavity is designed to keep the relative humidity under 10% if the water intake were five times this amount.

After glass seals are silver brazed into the ends of the copper tube of the repeater, the helium test is repeated to check the braze and to recheck the seal. For this test the entire repeater must necessarily be submerged in high-pressure helium. Obviously, to sense a possible passage of the gas from the outside to the inside, the leak-detector vacuum system must be connected to the internal volume of the repeater. For this and other reasons, a small-diameter tube that by-passes the seal is provided as a feature of the seal design. After the leak integrity of the repeater is established by this means for all but the access tube, this tube is then used as a means of vacuum drying the repeater and then filling it with extremely dry nitrogen. Following this, the tube is closed by welding and brazing. This closure is then the only remaining leak possibility and is checked by a radioisotope-leak test.

Radioisotope-Leak Test

Of various methods of detecting the passage of very small amounts of a liquid or a gas from the outside to the inside of a sealed repeater, a scheme using a gamma-emitting radioisotope appeared to be the most applicable.

The relatively small region of the welded tube referred to previously, is surrounded by a solution of a soluble salt of cesium-134. With the entire repeater in a pressure tank, hydraulic pressure in excess of service pressures is applied for about 60 hr. The repeater is removed from the tank, the radioactive solution is removed, and the test region is washed by a special process so as to be essentially free from external radioactivity. A special geiger counter is applied to the region. If there has been no leak, the gamma radiation reads a low value. If an intake has occurred of as much as one milligram of the isotope solution, the radiation count is about four to five times greater than that of the no-leak condition. The rate of leak indicated is an acceptable measure of the soundness of the repeater closure.

The helium and subsequent isotope-leak tests are made on a repeater when its glass seals are installed and are performed again on each rubber seal after it is brazed in place.

Electrical Tests

Prior to assembly into the repeater, the various networks are tested under conditions simulating as nearly as is feasible the actual operating conditions of the particular network. The input and output coupling networks and the beta networks enter directly into the insertion gain and hence are held to very close limits. To insure meeting these limits, elements which go into a particular network are

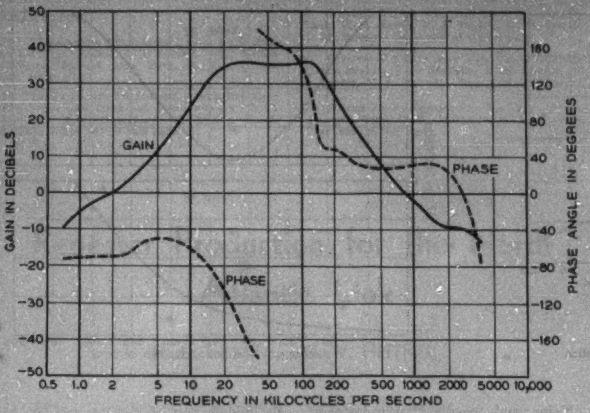


Fig. 18. Mu-Beta gain and phase

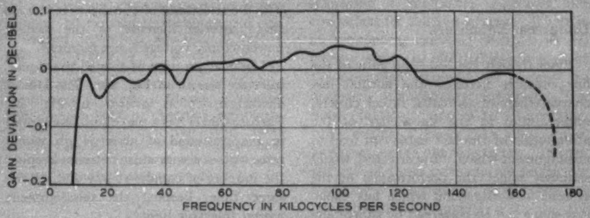


Fig. 19. Repeater deviation from 36.9 NM design cable

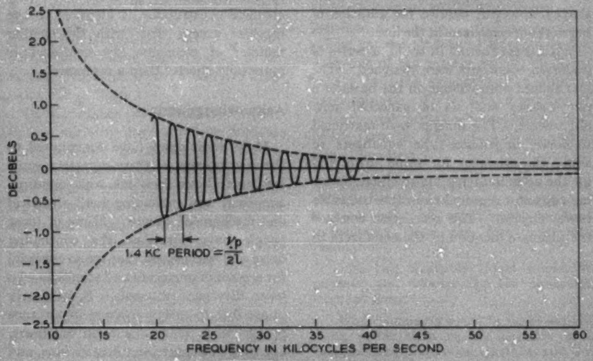


Fig. 20. Interaction ripple for TAC system

matched and adjusted as a group, before assembly into the network.

Repeater units are tested for transmission performance both before and after closing. These tests consist of: Mu-Beta measurements (simultaneous measurements of gain and phase of the feedback loop, noise, modulation, insertion gain at many frequencies, and exact frequency of the fault-location crystal and crystal-peak gain). Modulation and crystal-frequency measurements are made with the repeater energized at 225-ma cable current and also at 245 ma as a check on the ultimate performance of the whole system, initially and after aging.

Performance of Repeaters

The phase and gain characteristics of the feedback loop of the repeater are shown in Fig. 18. It will be noted that, at the upper edge of the band, the feedback is a little less than the 33- to 34-db set as the objective. Additional elements could have been used in the interstages to increase the feedback, but the return per element is small. Since any element is a potential hazard, the lower feedback is acceptable.

The deviation of the insertion gain of the repeater from the loss of 36.9 nautical miles of design cable at sea bottom is shown in Fig. 19. This is well within the objective of ±0.05 db.

It has been pointed out that the repeater input and output impedance do not match the cable impedance. This results in ripples in the system frequency characteristic due to reflections at the repeater. These are shown in Fig. 20.

The noise performance of the repeater is determined by the input tube and the voltage ratio of the input coupling network. Amplifier noise referred to the input is shown in Fig. 21. At the upper frequencies, the repeater contribution to cable noise is very small. At the lower frequencies, while the repeater noise is considerably greater than thermal noise, this does not degrade performance because of the lower cable attenuation at these frequencies.

Manufacturing Drawings

Because of the extraordinary nature of many of the manufacturing problems associated with undersea repeaters, it was determined at the outset that a so-called singled-drawing system would be used. For this reason, considerably more information is supplied than is normal. The effect is illustrated best in the rather large number of drawings that consist of text material outlining in detail a specific manufacturing technique. Such drawings specify the devices, supplies, and work materials needed to perform an operation, and the step-by-step procedure. Of course, these papers are by no means a substitute for manufacturing skill. Primarily, they insure the continuance of practices proved to be effective with the Havana–Key West project.

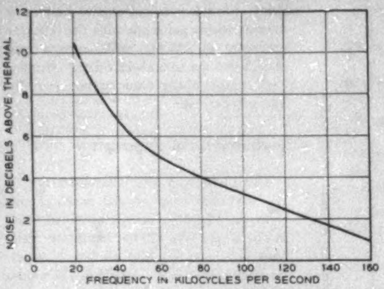


Fig. 21 (above). Repeater noise

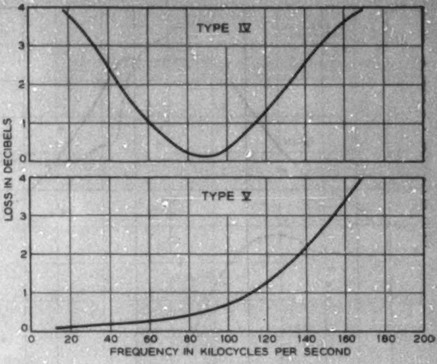


Fig. 23 (right). Equalizer loss characteristics

Repair Repeater

The repair repeater, used to offset the attenuation of the excess cable which must be added in making a repair, is basically the same general design as the line repeater. It employs a 2-stage amplifier, designed to match the loss of 5.3 nautical miles of cable to within ±0.25 db. The larger deviation, compared to the line repeater, is permissible since few repair repeaters are expected to be added in a cable. The input and output impedances match the cable. As in regular repeaters, a crystal and gas tube are provided for maintenance testing. The crystals give approximately 25-db increase in gain and are placed between 173.5 and 174.1 kc, so as not to duplicate any frequencies used in the line repeaters. The crystal-frequency spacing is 100 cycles.

Wherever possible, the same components and mechanical details are used in the repair repeaters as in the line repeaters. When changes in design were necessary, these were modifications in the existing designs rather than new types. Capacitors are like those of line repeaters. Except for the length of the container, the enclosure is identical to the line repeater.

Noise and overload considerations restrict the location of a repair repeater to the middle third of a repeater section.

Undersea Equalizers

Even though the insertion gain of the line repeater matches the normal loss characteristic of the cable rather closely, uncertainties in the knowledge of the attenuation of the laid cable can lead to misalignment which, if uncorrected, would seriously affect the performance of the system. Misalignment which has cable-loss shape can be corrected by shortening or lengthening the cable between repeaters, at intervals, as the cable is laid. Other shapes, however, require the addition of networks or equalizers in the line.

With these factors in mind, a series of undersea equalizers were designed. The loss shapes were chosen on the basis of a power-series analysis of expected misalignments. The designs were restricted to series impedance-type equalizers to avoid the necessity for shunt arms and for the accompanying high-voltage blocking capacitor required to isolate the cable power circuits. This restriction confines the ultimate location of the equalizers to the middle part of repeater sections to minimize the reaction of the poor repeater impedance on the equalizer characteristic. The d-c resistance of equalizers is low, so that material increase of the system power-supply voltage is not required.

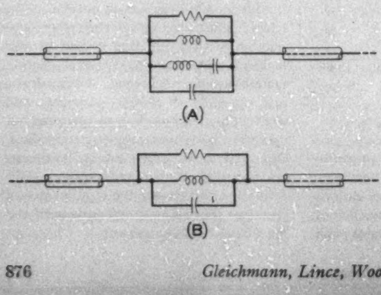


Fig. 22. A—Schematic of type-IV equalizer. B—Schematic of type-V equalizer

The configuration of two of the equalizers are shown in Fig. 22. The loss of characteristics is shown in Fig. 23. Each equalizer has a maximum loss spread in the passband of about 4 db which represents a compromise between keeping the number of equalizers low and, at the same time, keeping the misalignment within tolerable limits.

The components used are modifications of the repeater components. The mechanical construction is identical to the repeater except that, with the smaller number of elements, the container is materially shorter than a repeater.

Acknowledgments

Many individuals have contributed to the development of these repeaters, some leading to basic decisions, some creating, adapting, and perfecting both electrical and mechanical designs. Many of these people have furnished the continuing drive and enthusiasm that are so essential for a team of engineers and scientists who have divergent interests. It is nearly impossible to assign relative importance to the work of transmission engineers, apparatus designers, mathematicians, and research scientists in the fields of materials and processes. Equally difficult is any realistic appraisal of the work of all of the technical aides and shop personnel whose contributions are so significant to the final product. The authors therefore acknowledge this large volume of effort, without listing the many individuals by name.

References

1. A Submarine Telephone Cable with Submerged Repeaters, J. J. Gilbert. *Bell System Technical Journal*, New York, N. Y., vol. 30, Jan. 1951, p. 65.

2. Transatlantic Telephone Cable System—Planning and Over-all Performance, E. T. Mottram, R. J. Halsey, J. W. Emling, R. G. Griffith. Page 842 of this issue.

3. System Design for the North Atlantic Link, H. A. Lewis, R. S. Tucker, G. H. Lovell, J. M. Fraser. Page 850 of this issue.

4. Electron Tubes for the Transatlantic Cable System, J. O. McNally, G. H. Metson, E. A. Veazie, M. F. Holmes. Page 898 of this issue.

5. Power-Feed Equipment for the North Atlantic Link, G. W. Meszaros, H. H. Spencer. Page 890 of this issue.

6. Network Analysis and Feedback Amplifier Design (book), H. W. Bode. D. Van Nostrand Company, Inc., New York, N. Y.

7. Repeater Production for the North Atlantic Link, H. A. Lamb, W. W. Heffner. Page 877 of this issue.

8. Cable Design and Manufacture for the Transatlantic Submarine Cable System, A. W. Lebert, H. B. Fischer, M. C. Biskeborn. Page 907 of this issue.

Repeater Production for the North Atlantic Link

H. A. LAMB W. W. HEFFNER

Synopsis: Production of the submarine telephone cable repeaters, designed to have a minimum trouble-free life of 20 years, required many new and refined manufacturing procedures. Care in the selection and training of personnel, manufacturing environment, inspection, and testing, were of great importance in the successful attainment of the ultimate objective. Although quality of product has always been of major significance in Western Electric Company manufacture, building electronic equipment for use at the bottom of the ocean, where maintenance is impossible and replacement of apparatus extremely expensive, required unusual manufacturing methods.

Manufacturing Objective

LATE in 1952, the manufacture of flexible repeaters for the North Atlantic link of the transatlantic submarine telephone cable system was allocated to the Kearny Works of Western Electric Company.

In accordance with established practice in initiating radically new products and processes, production of these repeaters was assigned to the engineer of manufacture organization rather than to regular manufacture in the telephone apparatus shops. The job was to produce 122 36-channel carrier repeaters and 19 equalizers capable of operating satisfactorily at pressures up to 6,800 psi (pounds per square inch) on the ocean floor, with minimum maintenance, for a period of at least 20 years. Initial delivery of repeaters was required for March 1954, less than a year and a half after the project started.

General Philosophy

Quality has always been the prime consideration in producing apparatus and equipment for the Bell System. There is an economical breaking point, however, beyond which the return does not warrant the abnormal expenditures required to approach theoretical perfection. The same philosophy applies to all manufactured commodities, be they automobiles, airplanes, or telephone systems. In general, all of these products are physically available for preventive and corrective maintenance at nominal cost. In the case of electronic repeaters at the bottom of the ocean, maintenance is impossible and replacement would be extremely expensive.

The general philosophy adopted at the inception of the project was to build integrity into the product to the limit of practicability. To do this, a number of fundamental premises were established, which form the foundation of all operations involved:

1. Manufacturing environment would be provided which, in addition to furnishing a desirable place to work, could be kept scrupulously clean and free from contamination.
2. The best available talent would be screened and selected for the particular work involved.
3. Wage payments would be based on day work, rather than on an incentive plan basis, because production schedules and the complexity of the operations did not permit the high degree of standardization essential to effective wage incentive operation.
4. A sense of individual responsibility would be inculcated in every person on the job.
5. Training programs would be established to thoroughly prepare supervisors, operators, and inspectors for their respective assignments before any work was done on the project.
6. Inspection, on a 100% (per cent) basis, would be established at every point in the process which could, conceivably, contribute to or affect the integrity of the product.

Preparation for Manufacture

Manufacturing Location

It appeared desirable to set up manufacture in a location apart from the general manufacturing area. Experience gained to date has indicated that this was the correct approach, since it provided a number of advantages:

1. Administration has been greatly facilitated by having all necessary levels of supervision located in the immediate vicinity of the work.
2. It was necessary for the people on the job to acquire and maintain a new philosophy of perfection in product, rather than a high output at an "acceptable quality level." This was easier at a separate location, since only one philosophy was followed throughout the plant.
3. Engineering, production control, service, and maintenance organizations were located close to actual production and had no assignments other than the project.
4. The small plant, due to its semi-isolation, tends to produce a very closely knit organization and good teamwork.

A large number of manufacturing locations were examined and the one selected was a 1-story modern structure in Hillside, N. J., which provided a gross area of 43,700 square feet.

The entire plant was air conditioned; in most cases, the temperature was controlled to minimum 73 F (degrees Fahrenheit), maximum 77 F. The air was filtered through two mechanical filters and one electrostatic filter. Relative humidity was maintained at maximum 40% in all but one area, the capacitor-winding room, in which it was necessary to maintain maximum 20% humidity to avoid mechanical difficulty with capacitor paper. While most of the air was recirculated, the air from the cafeteria, cleaning room, locker, and toilet rooms was exhausted to the outside atmosphere. Two separate air-conditioning systems were in use. One, of 300-

---

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the AIEE Winter General Meeting, January 21–25, 1957.

H. A. Lamb and W. W. Heffner are with the Western Electric Company, Hillside, N. J., and Fairlawn, N. J., respectively.

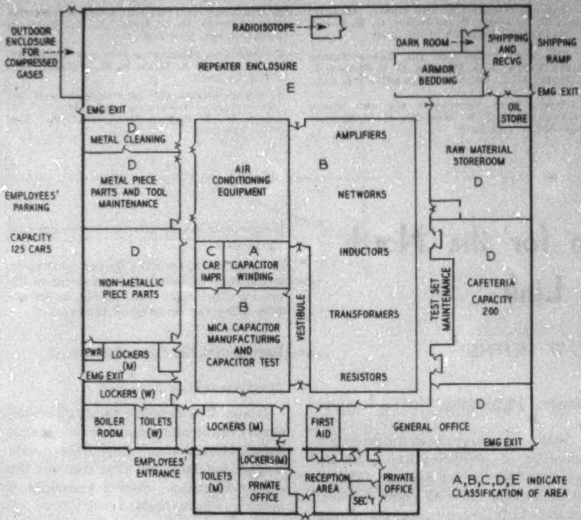

OUTDOOR ENCLOSURE FOR COMPRESSED GASES
RADIOISOTOPE
REPEATER ENCLOSURE
E
DARK ROOM
SHIPPING AND RECVG
SHIPPING RAMP
ARMOR BEDDING
EMG EXIT
OIL STORE
EMG EXIT
D
METAL CLEANING
D
METAL PIECE PARTS AND TOOL MAINTENANCE
AIR CONDITIONING EQUIPMENT
AMPLIFIERS
B
NETWORKS
RAW MATERIAL STOREROOM
D
EMPLOYEES' PARKING
CAPACITY 125 CARS
D
NON-METALLIC PIECE PARTS
C
CAP IMPR
A
CAPACITOR WINDING
B
MICA CAPACITOR MANUFACTURING AND CAPACITOR TEST
VESTIBULE
INDUCTORS
TRANSFORMERS
RESISTORS
TEST SET MAINTENANCE
D
CAFETERIA CAPACITY 200
PWR
LOCKERS (M)
EMG EXIT
LOCKERS (W)
BOILER ROOM
TOILETS (W)
LOCKERS (M)
FIRST AID
D
GENERAL OFFICE
EMG EXIT
EMPLOYEES' ENTRANCE
TOILETS (M)
LOCKERS(M)
PRIVATE OFFICE
RECEPTION AREA
SEC'Y
PRIVATE OFFICE
A,B,C,D,E INDICATE CLASSIFICATION OF AREA

Fig. 1. Plant layout

ton capacity, provided for most of the plant, while a smaller unit of 30-ton capacity served the capacitor winding, testing, and impregnating rooms. Each installation had its own air filtering and conditioning equipment.

PLANT LAYOUT

The plant layout is illustrated in Fig 1. All working areas, with the exception of the repeater enclosure area, were individually enclosed, and walls from approximately 4 feet above the floor were almost entirely of reinforced glass. This arrangement facilitated supervision by other than first-line supervisors, who were located with the groups, and provided a means of viewing the operations by the many visitors at Hillside, without contaminating the critical areas or disturbing the operators.

ANALYSIS OF DESIGN FOR FACILITIES AND OPERATIONS

In analyzing the design for manufacture there were, of course, numerous instances where conventional methods and facilities were entirely adequate for the job. Since their inclusion would contribute little to this article, the description will be confined to those cases which are new or unusual.

COLLABORATION WITH BELL TELEPHONE LABORATORIES IN PREPARATION OF MANUFACTURING INFORMATION

Early in 1953, a co-ordination committee was established, consisting of representatives from the various Laboratories design groups and Western Electric engineers, which met on a biweekly basis during the entire period preceding initial manufacturing operations. These meetings provided a clearing house for questions and policies of a general nature for this particular project and served to keep all concerned informed as to the progress of design and the preparations for manufacture.

It is customary, during the latter stages of development of any project at the Laboratories, for Western Electric engineers to participate in the preparation of manufacturing information as an aid in pointing the design toward the most economical and satisfactory production methods and facilities. Since the decision to use the Bell System repeater in the transatlantic system was based on the performance of the Key West–Havana installation, and the fact that changes in design would require further trials over an extended period, only minor changes to facilitate manufacture were made. Further, since some experience had been gained by the Laboratories in producing repeaters for that installation, it was decided to pool effort in preparing the manufacturing process information, which is normally Western Electric's responsibility. Close co-operation of the two groups therefore has resulted in the production of repeaters which are essentially replicas of those in the initial installation except for the internal changes necessary to increase transmission capacity from 24 to 36 channels.

OTHER WESTERN ELECTRIC LOCATIONS AND OUTSIDE SUPPLIERS

During the development work on the Key West–Havana repeaters, the Hawthorne Works of Western Electric had furnished the molybdenum-permalloy cores for certain inductors, the Tonawanda Plant had furnished mandrelated resistance wire, and the Allentown Plant had fabricated the glass seal subassemblies. Since the experience gained in this development work was extremely valuable in producing the additional material required for the transatlantic system and since the facilities for doing the work were largely available, these various locations were asked to furnish similar material for the project. Although the Kearny Crystal Shop had not been involved in the Key West–Havana project, arrangements were made to make the crystals there for this project, since facilities were available, and because of a considerable amount of experience in producing precision units.

SUBCONTRACTED OPERATIONS

While it was believed, initially, that all component parts for repeaters should be manufactured by Western Electric, critical analysis indicated that it was neither desirable nor economical in certain cases. One of the outstanding examples in this category is the hardened and ground chrome-molybdenum steel rings that constitute the strength members in the repeater and sustain the pressures developed on the ocean bottom. Purchasing the many large and varied machine tools and associated heat-treating equipment necessary to produce these parts would have required a substantial capital expenditure and additional manufacturing space. Arrangements were therefore made with a highly qualified and well-equipped supplier to produce the rings, of material furnished by Western Electric, which had been previously inspected and tested to very stringent requirements.

The situation attending the manufacture of a relatively small number of comparatively large copper parts used in the rubber and core tube seals was much the same. Here, again, the large-size machine tools and additional manufacturing space, required for only a short time, would have increased the over-all cost of the project considerably. These parts therefore were subcontracted in the local area and inspection was performed by Hillside inspectors.

A safeguard, in so far as integrity is concerned, was provided by the fact that these were individual parts that could be reinspected at the time of delivery. No subassembly operations that might possibly result in oversight of a defect, were subcontracted.

MANUFACTURING CONDITIONS

Two major problems confronted us in planning the manufacture of repeaters; first, to produce units that were essentially perfect and, second, to prevent the contamination of the product by any substance that might degrade its performance over a long period. In approaching both of these objectives, it was realized that the product had a definite economic value which the cost of production should not exceed. Therefore in many cases it was necessary to rely on judgment, backed by considerable manufacturing experience, in determining when the "point of no return" had been reached in refining processes and practices.

The initial approach to this phase of the job was to classify, with the collaboration of Bell Telephone Laboratories, all of the manufacturing operations involved as to the degree of cleanliness required. In setting up these criteria, it was necessary to evaluate the importance of contamination in each area and the practicability of eliminating it at the source, or to ensure that whatever foreign material might accumulate on the product would be removed.

A representative case is the machining of piece parts. While the shop area is cleaner, perhaps, than any similar area in industry, the very nature of the work is such that immediate contamination cannot be avoided, since material is being removed in the form of chips and turnings, and a water-soluble oil is used as a coolant. In this instance, however, the parts can be thoroughly cleaned and their condition observed before leaving the area. Conversely, in the case of an operation such as the assembly of paper capacitors in a container which is then hermetically sealed, it is vitally necessary to ensure that both the manufacturing area and the processes are free from, and not conducive to producing, particles of material which are capable of causing trouble.

The various classifications established for the production areas include specific requirements as to temperature, relative humidity, static pressure with respect to adjacent areas, cleanliness in terms of restrictions on smoking and the use of cosmetics and food, and the type and use of special clothing.

SPECIAL CLOTHING

Employees' clothing was considered one of the most important sources of contamination for two reasons; first, because of the foreign material that could be collected upon it and carried into the manufacturing areas and, second, because various types of textiles in popular use are subject to considerable raveling and fraying.

After considerable study of many types of clothing for use in critical areas, the material adopted was closely woven Orlon, which has proved to be acceptably lint-free. The complete uniform, supplied at no cost to the employees, consists of slacks and shirts for both male and female employees, Orlon surgeon's caps for the men and nylon-visored caps for the women. In addition, shoes, without toecap seams, were provided. Nylon smocks were furnished to protect the uniforms while employees moved from locker rooms to the entrance vestibule. Two changes of clothing were provided each week, and the laundering was done by an outside concern.

Employees to whom this special clothing was issued were paired for locker use. Each pair kept their uniforms and special shoes in one locker and their own clothes and shoes in the other. This prevented the transfer to the uniforms of any foreign material that might exist on the street clothing. At the entrance vestibule to the *A*, *B*, and *C* areas, Fig. 1, the employees were required to clean their shoes in the specially designed facilities provided and to wash their hands in the wash basins installed for this purpose. Smocks were then removed and hung on numbered hooks that line the walls at the end of the vestibule. Employees were then permitted to go to their work positions within the inner areas. At any time that it was necessary for employees to leave the work areas for any purpose, they were required to put on their smocks in the vestibule and upon their return, to go through the cleaning procedure again.

Employees in the other areas were provided only with smocks, mainly for the protection of their clothes, since the work involved could soil or stain them, but could not be contaminated from the clothing.

CLEANING

Schedules were established for cleaning the areas at regular intervals, the frequency and methods depending upon the type of manufacturing operations and the activity. Usually, the vinyl plastic floors were machine-scrubbed and vacuum-dried. Walls, windows, and ceilings were cleaned by hand with lint-free cloths. Manufacturing facilities such as bench tops, which were linoleum-covered, were washed daily. Test sets, cabinets, test chambers, and bench fixtures were also cleaned daily. Hand tools were cleaned at least once a week by scrubbing with a solution of green soap, rinsing in distilled water followed by alcohol, and then dried in an oven.

DUST COUNT

Since it was impossible to determine what contaminating material in the form of air-borne particles might be encountered from day to day, and what the effect might be during the life of the repeaters, the general approach to this problem was to control, so far as possible, the amount of dust within the plant.

To verify, continuously, the over-all effectiveness of the various preventive measures, dust counts were made in each classified area at daily intervals, using a Bausch and Lomb dust counter. This device combines, in one instrument, air-sampling means and a particle-counting microscope. Over a 2-year period, it has been possible to maintain, in certain areas, a maximum dust count of between 2,000 and 3,500 particles per cubic foot of air with a maximum size of 10 microns. Control checks, taken outside the building at the employees' entrance, generally run upward of 25,000 particles per cubic foot, a good portion of which are of comparatively large size.

Production and Personnel

Equipping the plant, obtaining and installing facilities, and selecting and training personnel proceeded on a basis of close overlapping, with receipt and analysis of Bell Telephone Laboratories' product design information. Because of the critical nature of the product, provisions were made not only for the most reliable commercially available utilities and services but also for emergency lighting service in some areas. It was necessary for maintenance and service staffs to be built up rapidly at the same time that the supervisory and manufacturing forces were being developed.

QUALIFICATION OF ALL PERSONNEL

Before employees were assigned to production work there were required to pass a qualification test established by the inspection organization to demonstrate satisfactory performance. Programs were therefore set up for "vestibule" training and qualification of new employees. This activity was carried on by full-time instructors who had been trained by Western Electric and Bell Laboratories engineers. Training was carried out in two stages:

1. (a) The employee received instruction and became acquainted with equipment and requirements. (b) A practice period was established in which the employee developed techniques and worked under actual operating conditions, with all work submitted to regular inspection.

2. During a qualification period, the employee was required to demonstrate that work satisfactory for project use could be produced.

The main objective during stage 1 was progressive quality improvement, and in stage 2 the maintenance of a satisfactory quality level over an extended period. Employees made a definite number of units at acceptable quality levels in order to qualify. The number of units required for training varied with the type of work and the ease with which it was mastered.

All personnel were required to pass qualification tests before being assigned to production work and were restricted to that work unless trained and qualified for other work. Employees trained on more than one job were requalified before being returned to a previous assignment.

Records of the performance of individual operators started in the training stage were continued after the employees were assigned to production work. The performance record of the operators was based on results obtained during the inspection of their work, while that of the inspectors was based on special-quality accuracy checks of their work.

PERSONNEL SELECTION

It was apparent that the new manufacturing techniques, including the cleanliness and quality demands, would necessitate that all shop supervisors and employees be very carefully selected. It also appeared (and this was subsequently confirmed) that after the careful selection and training of supervisors, long training periods would be required for specially selected shop employees.

In selecting first-line shop supervisors, such factors as adaptability, personality, and ability to work closely with the engineers were of paramount importance. For the parts and apparatus included in their responsibility, they were required to learn thoroughly the design, the operations to be performed, the facilities to be used, the data to be recorded, the cleanliness practices to be observed—and, in most cases, prepare themselves to be able to perform practically all of the operations, because subsequently they would train selected operators to perform critical operations to very high-quality standards under rigidly controlled manufacturing conditions. As shop supervisors and employees were assigned to the manufacture of repeaters, they were thoroughly indoctrinated in the design intent and the new philosophy of manufacture.

Standard ability and adaptability tests were used in a large number of cases to assist in proper selection and placement of technicians: tests for finger and hand dexterity; for sustained attention; eye tests, including perception and observation; and tests for reaction time of the right foot after a visual stimulus. (The latter test was relatively important for induction-brazing operations.) Other requisite considerations were a high degree of a dependability and integrity, involving intellectual honesty and conscientious convictions; capability of performing tedious, frustrating, and exasperating operations against ultrahigh-quality standards, verifying their own work; perseverance and capability to adapt easily to changes in assignment and occupation or the introduction of design changes. We considered whether or not the employees would stand up under "fishbowl" operations, wherein they would receive a considerable amount of observation from high levels of Western Electric Company and Bell System management, and other visitors. Also, could they duplicate high quality frequently after qualifying for a particular operation?

During the period of repeater manufacture, the number of employees rose from less than 50, in January 1954, to a maximum of 304 by February 1955, after which there was a gradual reduction to a level of about 265 employees for 6 months and then a gradual falling off as we were completing the last of the project. In the period from May to December 1954, between 30 and 45 employees were constantly in training prior to being placed on productive work. During 1955, this decreased to practically no employees in training during the mid-part of the year and thereafter training was required merely to compensate for a small labor turnover and employee reassignment. It is significant that labor turnover was very low and attendance was exceptionally good during the life of the Hillside operations.

PERSONNEL TRAINING

The original plan, which was generally followed, was to prove in the tools for each phase of the job, followed by an intensive program of training. Indoctrination of laboratory technicians could be considered as "vestibule training" in that they were acclimated to the area and conditions, given oral instruction in the work, then given practice materials and demonstrations and, when qualified, were started on making project material. To do this, extra supervisors were required at the beginning of the job. A supervisor trained a few employees, qualified some of them, and began work on the project. Another supervisor was then required to train additional employees who, as they became qualified, were transferred to the supervisor responsible for making project apparatus. Additional testing of the employees, instruction and reinstruction and, in some cases, retraining were required. In practically all cases, we were able to fit an employee selected for work at Hillside into some particular group of operations. The extra emphasis on selection and training created a well-balanced team that later resulted in considerable flexibility. During all of this training our supervisors worked closely with engineers and inspectors who understood the design intent and the degree of perfection required.

At the beginning, each technician was trained for only one operation of a particular job, such as: 1. winding type-X capacitors, 2. impregnating all paper capacitors, or 3. winding type-Y transformers, and so became an expert on this one operation. Later, the tours of duty for many technicians were broadened to cover several operations.

COMMUNICATIONS

To keep employees informed, the entire group was occasionally assembled, and informative talks were presented on current production plans and future business prospects. Motion pictures were shown of the cable-laying ships and the operations of cable splicing and cable laying. A display board, showing all of the repeater components, was mounted on the wall of the cafeteria. This informed the operators exactly where the parts were used in apparatus; also where their products went into the wired repeater unit, and how all electric apparatus was enclosed against sea pressure in the final repeater. In small groups, all of the employees at Hillside were given a short guided tour of the plant to see the facilities and hear a description of the operations being performed in each area. These communications were extended to everyone at the Hillside Plant, including those who did not work directly on the product. It was our conviction that the maintenance men, boiler operators, oilers, station wagon chauffeur, janitors, and clerical workers in the office were all interested and could do a better job if kept informed of the needs and progress of the project.

SCHEDULING

Capacity was provided at the Hillside Shop to manufacture a maximum of 14 repeaters in a calendar month. This envisioned 6-day operation with some second-shift and third-shift operations; due allowance was made for holidays and vacations, so that the annual rate would be approximately 160 enclosures per year. (An enclosure is either a repeater or an equalizer.)

Some of the facilities and raw materials were ordered late in 1953. This ordering expanded early in 1954 and continued through 1955 to include parts to be made by outside suppliers and the parts and apparatus to be made at Hillside. Apparatus designs were not all available at the beginning of the job, and the ultimate quantities required were also subject to sharp change as the project developed, thus further complicating the scheduling problem.

Because of the time and economic factors involved, coupled with the developmental nature of the product and processes, one of the most difficult and continuing problems was the balancing of production to meet schedules. For this task, "tree charts" were devised for the apparatus codes and time intervals in each type of repeater or equalizer for each project. Each chart was established from estimates of the time required to accomplish the specified operations and the percentage of good product each major group of operations was expected to produce.

Raw Materials

Many of the specifications were written around the specific needs of the job and embodied requirements that were considerably more stringent than those imposed on similar materials for commercial use. As a result, it was necessary for many suppliers to refine their processes and, in some cases, to produce the material on a laboratory basis.

One example is the container, or repeater enclosure, which consists, in part, of a seamless copper tube approximately $1^{3}/_{4}$ inches in diameter having a 1/32-inch wall and which is approximately 8 feet long. This material was purchased in standard lengths of 10 feet. The basic material was required to be phosphorus deoxidized copper of 99.80% purity. The tubing, as delivered, had to be smooth, bright, and free from dirt, grease, oxides (or other inclusions including copper chips), scale, voids, laps, and slivers. Dents, pits, scratches, and other mechanical defects could not be greater than 0.003 inch in depth. The tubing had to be concentric within 0.002 inch and the curvature in a 10-foot length could not exceed $^{1}/_{2}$ inch, to facilitate assembly over the steel rings.

Only one supplier was willing to accept orders for the tubes, and only on the basis of meeting the mechanical requirements on the outside surface. To establish a source of supply, it was necessary to accept the supplier's proposal on the basis that some of the tubes produced could be expected to meet requirements on the inside as well as the outside surface. Inspection of the inside surface was performed with a 10-foot borescope.

The supplier then set aside, overhauled, and cleaned a complete group of drawing facilities for this project. In addition, a number of refinements were made in lubrication and systematic maintenance of tools. After all refinements were made and precautions taken, however, the yield of good tubes in the first 400 produced was less than 1.0%. Consultations with Western and Bell Laboratories' engineers, and the supplier's cooperation, raised the yield to approximately 50%.

Procurement of satisfactory mica laminations for capacitors introduced an unusual problem. The best grade of mica available in the world market was purchased, which the supplier, under special plant conditions, split and processed into laminations. Despite care in selection and processing, only 50% of the 250,000 laminations purchased met the extremely rigid requirements for microscopic inclusions and delaminations, and less than 8% survived the capacitor manufacturing processes.

A large number of the parts, and the most complex, are made from methyl-methacrylate (Plexiglass). At the time manufacture began, there was little, if any, experience or information available on machining this material to the required close tolerances and surface finish. Consequently, considerable pioneering effort was expended in this field before satisfactory results were obtained.

The methacrylate parts cover a wide range of size and complexity, from tubular housing $1^{1}/_{8}$ inches in diameter and $4^{7}/_{8}$ inches long to tiny spools 1/8 inch in diameter and 1/16 inch long. Most of the parts are cylindrical in shape with some semicylindrical sections that must mate with other sections to form complete cylinders. Others have thin fins, walls, flanges, and projections. Five representative parts are shown in Fig. 2.

Methyl-methacrylate has a tendency to chip if tools are not kept sharp and if care is not used in entry or exit of the tool in the work, particularly in milling. In some cases it is necessary, with end milling, to work the cutter around the periphery of the area for a slight depth so that subsequent cuts will not break out at an unsupported area. Normally, with a sharp cutter and a 0.010-inch finish cut, and a slow feed, chipping will not result. High-speed steel tools with zero rake were used for turning and boring operations. Standard high-speed milling cutters and end mills were used for milling except for the cutting edges, which are honed to a fine finish. A clearance angle of 7 degrees for milling and 10 to 15 degrees for lathe work was found most satisfactory. In lathe work, the general rule was light feeds (0.003 to 0.005 inch) and small depth of cut. However, the depth of cut could be safely varied over a wide range, depending upon many factors, such as type of part, quality of finish, machine and tool rigidity, effective application of coolant, and tooling to support and clamp the part. In one operation of boring a blind hole, $1^{3}/_{16}$ inches in diameter and $4^{1}/_{8}$ inches deep, within ±0.002 inch, the boring terminates in simultaneously facing the bottom of the hole square with its axis. A cut 19/32 inch deep with a light feed was taken with a specially designed boring tool and with the coolant being fed through the shank to the cutting edge. All completely machined parts were annealed for 12 hours at 175 F.

Highlights in Assembly and Brazing

Repeater units are encased in hardened steel rings which previously had been tested at 10,000-psi hydraulic pressure. This pressure is approximately 50% higher than the greatest pressure expected at ocean bottom. The steel rings were encased in a copper sheath and closed at each end with a glass-to-Kovar seal, with the central conductor coming through the glass to the outside. The copper sheath was then shrunk to the steel rings and glass seals using 6,000-psi hydraulic pressure, and the glass seal was then high-frequency brazed to the copper sheath.

To keep the ocean bottom pressure off the glass seals and also to terminate the cable insulation, a rubber seal is brazed in to the copper container tube adjacent to each glass seal. This rubber seal consists of rubber bonded to brass, which has been brazed to the copper portion of the seal. The rubber terminates in polyethylene through five steps of compounds containing successively less rubber and more polyethylene. The polyethylene can be readily bonded by molding to the polyethylene insulation of the cable. The central conductor passes through a central brass tube in the rubber seal, which is also bonded to the rubber.

To protect the rubber seals from the deleterious effects of salt-water immersion for long periods, a copper-core tube is brazed over each rubber seal. The core is arranged to equalize the pressure inside and out when submerged at ocean bottom pressure. This is accomplished with a bulge of neoprene filled with polyisobutylene, on the far end of the core tube, which transmits the pressure to the inside of the core tube seal.

To make doubly sure that no salt water reaches the rubber seal, a copper cover is brazed into the container outside the core tube connector on each end. This cover is also brazed to the core tube connector. The interstice between each of these four seals is filled with polyisobutylene, which is viscous and inert and has very good insulating qualities.

Each end of the repeater closure, Fig. 3, contains five successive brazed joints. Any one of these ten brazes, if not perfect, could cause the loss of the repeater closure and jeopardize the entire repeater. All of these brazes were made with the repeater in a vertical position to ensure an even distribution of the brazing alloy fillet around the joint.

An up-ending device was provided at the pit-brazing location to raise to repeater on its carrier to a vertical position with either end up, and move it into position for brazing. The repeaters were brought into the brazing area on an overhead monorail and an electric hoist. The shorter repeater assemblies, before core tube and cable stub assembly, were up-ended by hand and brazed from a raised platform.

It was necessary to make all of these brazes by high-frequency induction heating, since the heat must be very intense, contained within a very narrow band, evenly distributed, and the area protected from oxidation by a somewhat reducing atmosphere. The heat had to be very intense since the time interval for the shortest braze was 10 seconds maximum and the longest 30 seconds. A large part of the heat was dissipated by being conducted at a high rate from the copper parts to the water in the cooling jackets used to contain the heat in a very narrow band.

Circulating cooling water within a jacket prevented heat from being conducted down the copper container tube to the preceding seals or to the repeater unit. This water-cooled jacket was positioned only 3/8 inch below the inductor, and the water was in intimate contact with the container tube, which is sealed off at both ends with rubber O-rings. In addition, for the glass seal braze, the glass inside the seal cavity was kept covered with water during the heat cycle. The water was fed in and siphoned out to a constant level which was kept under observation by the operator and the inspector to make sure that the glass was covered at all times. The rubber seal was also water-jacketed on the inside of the seal to prevent deleterious effects of the heat on the rubber insulation around the central conductor. The inner cover braze was quenched before the 10-second maximum interval had expired to ensure that the heat did not penetrate to the polyisobutylene at a sufficient rate to deteriorate it or the rubber inside.

Distribution of the heat around the container tube at the braze area was controlled by locating the work in the inductor so that the color came up essentially evenly all the way around and at the proper level to bring a fillet up to the top of the braze joint within the allowable time limit. The time limit was determined by experiment so that none of the previously assembled parts were damaged by the heat. This determination of the proper heat pattern and the prevention of overheating required the development of considerable skill on the part of the operator. The variables encountered made it essential to rely on an operator to control the heat rather than to utilize the timer with which the induction heating equipment is normally controlled.

The area to be heated for brazing was protected from oxidation by enclosing it in a separable transparent plastic box and flooding the interior with a gas consisting of 15% hydrogen and 85% nitrogen. This atmosphere is somewhat reducing and not explosive. The brazing surfaces of the parts were chemically cleaned immediately before assembly and extreme care was exercised to keep them clean until brazed.

The container tube was shrunk to the respective glass, rubber, core tube, and cover seals with the use of hydraulic pressure so that the surfaces to be brazed and the brazing alloy were in intimate contact within the brazing area. If the parts were clean and kept from oxidizing by the protective atmosphere, the alloy would flow upward by capillary action and form a fillet around the top of the seal, impervious to any leak.

The braze in each case was then leak-tested with a helium mass-spectrometer type of leak detector. A gas pressure of helium at least 25% greater than the maximum pressure to be encountered at ocean bottom was used. In addition, a radioisotope was used to test the effectiveness of the final tubulation pinch welds and overbrazes which were kept open for the leak tests under high-pressure helium. These tests were made with water pressure about 25% greater than the maximum ocean bottom pressure.

The completed repeater was inserted in a chamber 80 feet long; the chamber was then filled with water and the pressure raised to 7,500 psi and held at that pressure for at least 15 hours. At the end of this period the closure had to show no sign of crushing or leaking.

The repeater unit sealed in the closure must be extremely dry to function properly. Any water vapor which might remain after the closure is sealed, or enter during the estimated 20-year minimum life, must be scavenged. A sealed desiccator with a thin diaphragm was therefore assembled into the repeater unit sections. After completely drying and sealing the repeater unit except for one tubulation, the diaphragm of the desiccator was ruptured by dry nitrogen pressure and, with the enclosure filled with dry nitrogen, the final tubulation was immediately sealed off. To ensure that the diaphragm was actually broken, a microphone was strapped to the outside of the repeater over the location of the desiccator and a second microphone arranged at the end of the closure to pick up background noises. A pen recorder was used to record the sound from the two microphones and also the change in nitrogen pressure. Three simultaneous pips on the chart gave definite indication that the diaphragm had ruptured and that the desiccant had been exposed to the internal atmosphere of the repeater.

Quartz Crystal Units Manufactured at Kearny

The primary purpose of the crystal unit is to provide the means of identifying and measuring the gain of each repeater in the cable. This basic crystal design is in common usage. The exacting specifications for this application, however, imposed many problems and deviations from normal crystal-manufacturing processes.

Raw quartz was specially selected for this crystal unit. The manufacturing process of reducing the quartz to the final plate followed the recognized methods through the roughing operations. Because of the rigid end requirements, the finishing operations were performed under laboratory conditions. Angular tolerances were one third of normal limits. No evidence of surface scratches, chipped edges, or other surface imperfections visible under 30× magnification were permitted. This resulted in a process shrinkage five times that experienced in normal crystal plate manufacture.

In this use, the crystal units were required to meet performance tests at currents as low as 1/1,000 microampere, far below the current values usually encountered. Improved soldering techniques had to be developed for soldering the gold-plated phosphor bronze and nickel wires used, because it was found that the electrical performance of the units was directly related to the quality of soldered connections.

Although one seventh of Western Electric's production of quartz crystal units are in glass enclosures, the applicable techniques in glass working required a complete revision. Glass components such as the stem and bulb purchased from established sources were found to be far below the standard required for this crystal unit. For example, the supplier of the glass tubing used in the manufacture of stems was required to meet raw material specifications that embodied coefficient of thermal expansion, softening point of glass, density, refractive index, and volume resistivity. The glass stems made from this tubing by regular manufacturers were found unacceptable and the processes used by these sources could not be readily adapted to meet the desired specifications. The glass stems contained four lead wires made from 30-mil grade *A* nickel wire butt welded to 16-mil light borated Dumet wire. To assure the quality of the metal to glass seal, each wire was inspected under 30× magnification for tool marks and other surface imperfections. The finished stem assemblies were inspected under 30× magnification for dimensions, workmanship, cleanliness, and minute glass imperfections, then individually stored in a sealed plastic envelope.

The glass bulb in this crystal unit is known as the *T921* design commonly used in the electron tube industry. The high quality required, however, made 100% inspection necessary. Examination under 30× magnification resulted in rejected bulbs for presence of scratches, open bubbles, chips, and stones. Physical limits for inside and outside diameters as well as wall thickness were causes for additional rejects. Only 1% of the commercial bulbs were found acceptable, and these were all stored in a sealed plastic envelope.

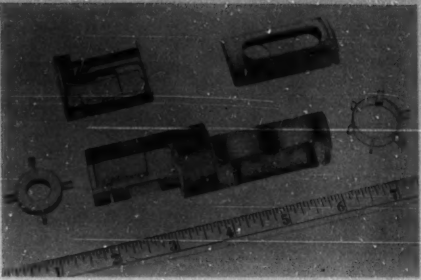


Fig. 2. Methyl methacrylate parts

The final major assembly operation consisted of sealing the glass bulb to the stem which had had the crystal subassembly welded to the nickel wires. The techniques for "sealing in" used in quartz crystal or electron tube manufacture were unsuited. Two important factors in this crystal unit, which required the development of new processes, were the proximity of soft-soldered connections to the sealing fires and the demands that the glass seal contain a minimum of residual tensile stress. These two problems were resolved collectively by performing the sealing operation on a single-spindle glass-sealing machine. Accurate positioning of the glassware and sealing fires, together with precise timing and temperature controls, achieved the desired results.

Evaluation of residual stresses were made by inspections with the use of a polarimeter, and by a thermal shock test. The maximum safe stress was established at 1.74 kilograms per square millimeter. The thermal shock test required successive immersion of the unit in boiling water and ice water. The electrical characteristics of these units exceeded all others made previously by Western Electric. The ratio of reactance to effective resistance Q was greater than 175,000, twice that ever previously produced and 17 times that required in the average filter crystal.

Stability for frequency and resistance was assured by a 28-day aging test. During this period, precise daily resonant frequency and resistance measurements were recorded against temperature within 0.1 degree centigrade. The maximum permissible change was 0.0005% in frequency and +5% to −10% in resistance.

Glass Seals Manufactured at Allentown

The glass seal used to close each end of the container for the repeaters and equalizers is manufactured at the Allentown Works of the Western Electric Company.

The unit is essentially a glass-bead-type seal. It insulates the central conductor of the repeater from the container and serves as a final vapor barrier between the cable and the interior of the repeater. As such, it backs up several other rubber and plastic barriers as shown in Fig. 3.

Fig. 4 shows the various components, subassemblies, and a cross section of the unit. The unit consists of the basic seal brazed in the Kovar outer shell, to which is brazed a copper extension provided with two brazing-ring grooves. One of these grooves is used in brazing the seal, along with support members, into a length of container tubing in the same manner as the seal is ultimately brazed into the repeater. Packaging of the seal in this manner was necessary to pressure test the seal. Under test, in a specially constructed chamber 10,000 psi of helium gas pressure was applied to the external areas of the packaged glass seal and a mass spectrometer-type leak detector was connected through the tubulation to the internal cavity of the packaged unit. In this manner, the interface of the glass to metal seal, the brazed joints, and the porosity of the metal were checked for leakage. The unit is left in this package for delivery to provide protection during shipment. Before the seal could be used, it was machined from the package by cut-

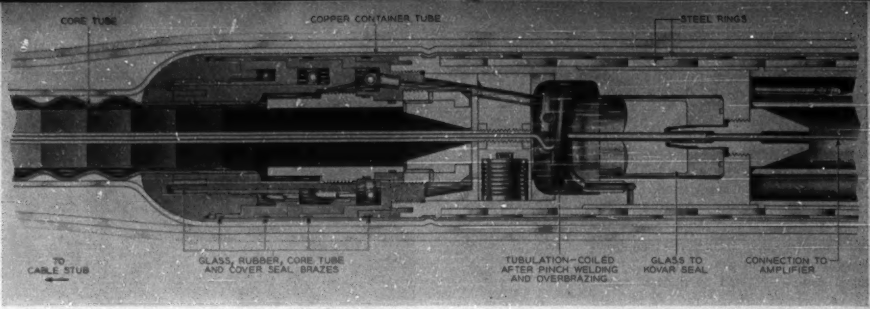

CORE TUBE
COPPER CONTAINER TUBE
STEEL RINGS
TO CABLE STUB
GLASS, RUBBER, CORE TUBE AND COVER SEAL BRAZES
TUBULATION—COILED AFTER PINCH WELDING AND OVERBRAZING
GLASS TO KOVAR SEAL
CONNECTION TO AMPLIFIER

Fig. 3. Repeater seals

ting the copper extension to length leaving the second groove for use in brazing the seal to the repeater and removing the container tubing and support members.

The basic seal consists of the cup, central conductor, and glass. The cup (the smaller cylindrical item in the upper left-hand corner of Fig. 4) was machined from Kovar rod. The wall of the cup is tapered from a thickness of 0.025 inch at the base to 0.002 inch at the lip. The last 0.006 inch of the lip is further tapered from this 0.002 inch to a razor edge. The internal surface is better than a 63-micro-inch turned finish and was also liquid-honed to give it a uniform matte finish. The central conductor (the slim piece in the upper right-hand corner of Fig. 4) was also machined from Kovar rod. Both the cup and central conductor were further processed by pickling, hypersonically cleaning in deionized water, and decarburizing. The glass, a borosilicate type of optical quality, was cut from heavy-walled tubing. The glass tubing was hand-polished, lapped, and etched to remove surface scratches, and to arrive at the specified weight. It was also fire-polished and hypersonically cleaned to remove all traces of surface imperfections and to assure maximum cleanliness.

To make the basic glass seal, the metal parts had to be oxidized under precisely controlled conditions. For the oxidizing operation, a suitable fixture was loaded with brazed shell-cup assemblies, central conductor assemblies, and a Kovar disk, which had been prepared in precisely the same manner as the cups and central conductors. The disk was carefully weighed before and after oxidizing and the increase in weight divided by the area involved yielded the weight gain due to oxidation for each run. Limits of 1.5 to 2.5 milligrams per square inch of oxide were set. This operation was performed by placing the loaded, sealed retort, through which passed a metered flow of dried air, into a furnace for a specified time-temperature cycle.

In the glassing operation the oxidized shell assembly, the carbon mold, and the central conductor were placed in a fixture and held in the proper relationship. The carbon mold served to support the glass while it was being melted, in that section between the cup and central conductor where the glass was normally unsupported. The prepared cut-glass tubing was loaded into the Kovar cup and the fixture was sealed into the retort. During the glassing cycle, a constant flow of nitrogen passed through the retort to provide an atmosphere which minimized any reduction or further oxidation of the already carefully oxidized parts. After the proper purging period, the retort was placed in the furnace. In the furnace, the glass melted and formed a bond with the oxidized Kovar of the cup and central conductor to form the seal. After the specified temperature-time cycle, the retort was removed from the furnace, allowed to cool partially, and then placed into an annealing oven.

Vertical furnaces and retorts were used for brazing, decarburizing, oxidizing and glassing. By varying the type of gases flowing into the retorts, atmospheres which are reducing, oxidizing, or neutral were obtained. To provide maximum uniformity of process, separate retorts and holding fixtures were provided for operations involving hydrogen and for air-nitrogen operations, so that a retort or a fixture used for hydrogen treatments was never used for oxidizing or glassing.

Pilot and Regular Production

We called our first efforts Practice Parts and Training; the next we called Pilot Production. Next, certain items identified as "trial-laying repeaters and oscillators" were manufactured for use in proving in the ship-laying gear. To prove in manufacturing facilities, a few unequipped housings were made without the usual electrical components normally in a repeater. Similarly, each of the apparatus components and parts required exploratory and pilot effort before regular production could be undertaken.

As might be expected, the manufacturing yield of components meeting all requirements was very low during the early stages of the undertaking. However, substantial improvement was brought about as experience was gained. Comments on some of the production problems, highlights, and yield results, are in the following paragraphs.

PAPER CAPACITORS

Paper capacitors were manufactured only after painstaking qualifying trials and tests had been performed on each individual roll of paper. Cycling and life testing, procurement of acceptable ceramic parts and gold-plated tape and cans, selection and matching of rolls of paper for winding characteristics, and similar problems, all had to be completely resolved to a point of refinement previously unattempted for telephone apparatus.

Composite percentage yield for all operations on paper capacitors is shown in Fig. 5. Yield is shown as the ratio of the finished units of acceptable quality to the number of units started in manufacture.

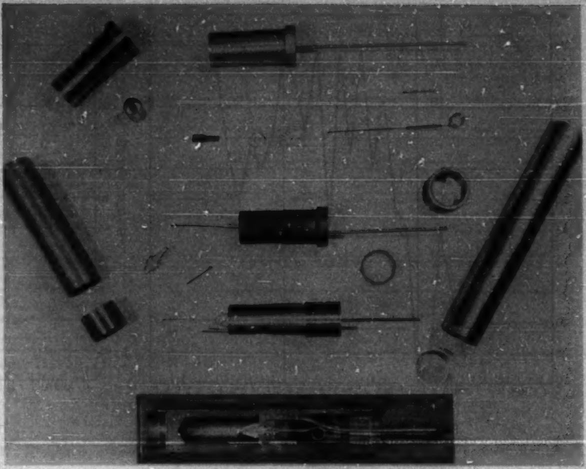


Fig. 4. Glass to Kovar seals

Mica Capacitors

Mica capacitors were made from only the most meticulously selected laminations, as mentioned earlier. Even the best mica is particularly susceptible to damage in processing. In spite of experience and knowledge of this, the multiple handling of the laminations contributed an unusually high material shrinkage as each separate lamination needed to be cleaned, then handled individually many times through the processes. The art of silk screening was applied to deposit silver paste in a specific area or areas on each side of a lamination. A sharply defined rectangular area was required so that when the laminations were superimposed one over another the desired capacitance would be obtained. Cementing of mica laminations onto machined methacrylate forms presented some additional problems through the bowing of the mica laminations as the cement cured. Obtaining screens that would give the proper length and width dimensions for the coated area, was another problem. A silk screen woven of strands of silk obviously limits, by the diameter of the threads, the extent to which the dimensions of an opening may be increased or decreased. Soft copper U-shaped terminals were used to clamp the layers of mica together into a stack. Control of the pressure used in crimping these terminals was found to be very critical in view of the exceptionally tight limits on capacitance and stability. Fig. 6 shows the composite yield at various times for all mica capacitors.

Resistors

There were three designs of ceramic resistors, which were resistance-wire wound on ceramic spools. These were intended to be assembled into the hole inside the core tube on which the paper capacitors were wound. Special winding machines equipped with binocular attachments were necessary to wind these resistors. Other resistors were hand wound on methylmethacrylate forms, or on the outside of the ceramic containers, for certain types of paper capacitors. Rough adjustments were required of the lengths of resistance wire prior to winding, and close adjustments to resistance values were made after the windings were completed and before leads were attached to resistors. Again it was necessary to provide periodic samples that could be placed on life test by the Laboratories to ascertain that the manufacturing processes were under control. These samples, in all possible cases, were taken from product that would normally be rejected because of some minor defect, but which would not in any way detract from the validity of the life tests. The making of hard solder splices between nichrome resistance wire and gold-plated copper leads, and keeping ceramic parts from coming in contact with metal surfaces and thereby being contaminated because of the ceramic's abrasive characteristics, were two major problems on resistors. Fig. 7 indicates resistor yields.

Inductors

Inductors comprised 20 different designs, most of which were air core, but there were some for which it was necessary to cement permalloy dust cores into pockets of the methacrylate form, and thereafter using wire on a shuttle, wind by hand the turns required to produce an inductor. These varied from a very small inductor, smaller in diameter than a pencil, to a fairly large "figure eight" inductor with turns having a major diameter of about $1^{1}/_{4}$ inches. Each layer of a winding was inspected with a microscope to insure that the wire had not been twisted or kinked, or that the insulation was damaged or uneven. Some of the shuttles became fairly long so that they could hold the amount of wire required to make a continuous winding. The operator's handling of this shuttle, as she moved it down around the openings in the methacrylate part, or placed it on a bench to proceed with the interleaving tape, demanded considerable dexterity and concentration to ensure that the shuttle was not turned over, which in effect would put a twist in the wire. Although best known means were used to sort cores for their magnetic properties prior to the time a winding was made, the limits on the inductors themselves were so close that subsequently a large number of windings were lost. The best cores that could be selected, plus the best winding practice, could not produce 100% of the inductors within the required limits. Crazing of the insulation on the wire; cementing together of two methacrylate parts or of permalloy cores into pockets of methacrylate parts, and handling those inductors having long delicate leads, were the most troublesome items on this apparatus. Fig. 8 shows manufacturing yield for inductors.

Networks combined several codes of component apparatus, such as a mica and a paper capacitor, resistor, and an inductor. Six networks were used in each repeater unit consisting of two interstage networks, an input, an output, and two beta networks. They demanded a most delicate wiring job, in that stranded gold-plated copper wires had to be joined in a small pocket in methyl methacrylate, where the least heat can be applied; otherwise the methacrylate is affected. After soldering, a minimum amount of movement of the stranded wire was permitted, inasmuch as the soldered gold-plated copper wire becomes quite brittle.

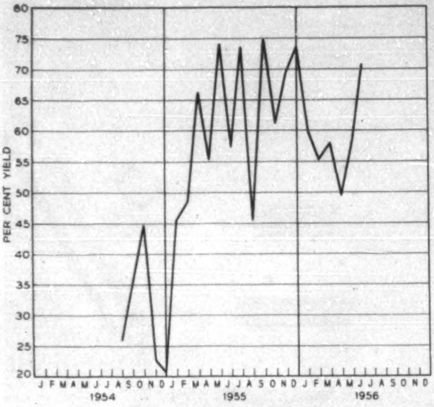


Fig. 5. Paper capacitor yield

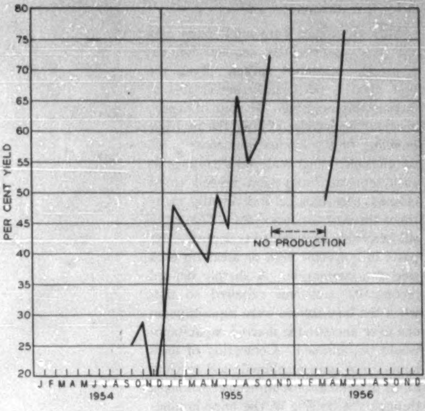


Fig. 6. Mica capacitor yield

REPEATER UNITS

Repeater units are wired assemblies consisting of 17 sections in which there are six networks, three electron tubes, one gas tube, one crystal, three high-voltage capacitors, one dessicator and two terminal sections. The successive build-up of these materials left little chance to make a repair because a splice in a lead was not permissible. It is during this assembly stage that a repeater receives its individual identity because of the frequency of the particular crystal assembled into the unit. A manufacturing yield of 100% was achieved in the assembly and wiring of repeater units.

It was necessary to calibrate the test equipment for this job very closely. Bell Telephone Laboratories and Western Electric worked at length to calibrate the testing details and the test sets for individual networks. Adjustments in components apparatus to bring the network to the fine tolerances required were accomplished by minute scraping of the silvered mica on a mica capacitor or removing turns from wire-wound inductors. The cementing of methacrylate parts, which was a troublesome item on mica capacitors and inductors, also had to be contended with on networks.

Packing and Shipping Co-ordination

Repeaters were packed in Western Electric specially designed 34-foot-long aluminum containers, weighing 1,000 pounds. Forty of these containers were made by an outside firm. Fig. 9 shows two containers tied down in a truck trailer. The repeaters were nested in a pocket of polyethylene bags containing shaped rubberized hair sections in order to cushion them during their subsequent handling and transportation. The instrumentation required with each case was tested, properly set, and inspected prior to its use on each outgoing case. The instruments were a shock recorder to register shocks in three planes, and a thermometer to register the minimum and maximum temperatures to which the repeater had been exposed. Arrangements were made with a commercial trucking company to provide three specially equipped truck trailers, which could be cooled by dry ice during hot weather and warmed by burning bottled gas during cold weather so as to control temperature within the 20 to 120 F called for in the repeater specification.

Appointment of a shipping co-ordinator supervisor added tremendously to the smooth functioning of services and provided the continuing vigilance required to protect repeaters and deliver them to the right place at the right time. His responsibility was to co-ordinate all the shipping information and arrangements from the time the item was ready for packing at the Hillside plant, through all trucking arrangements to the armoring factory, to the airport, to England, and to follow, with statistical data and reports, each enclosure up to the point at which it was possible to record the date on which the repeater was laid or stored in a depot.

Inspection Plan and Procedures

GENERAL

It is axiomatic that quality is not obtained by inspection but must be built into the product. However, the Inspection Organization does have the responsibility of certifying that the desired quality exists. Our evaluation indicated that the ordinary inspection screening would be inadequate to ensure the high degree of integrity demanded and that additional safeguards would have to be provided. These controls were achieved, in a practical way, by:

1. Selective placement, intensive training, and subsequent qualification testing of all personnel.
2. Inspection during manufacturing operations in addition to inspection of product after completion, and regulating inspection so that critical characteristics received repetitive examination during the process of manufacture and assembly.
3. A maintenance program for inspection and testing facilities which provided checks at considerably shorter intervals than is considered normal.
4. Inspection and operating records and reports that point out areas for corrective measures.
5. Records of quality accuracy for all inspection personnel as an aid in maintaining the high quality level.
6. Verification of all data covering process and final inspection as a certification of the accuracy of these data and certification that the apparatus satisfactorily meets all requirements.

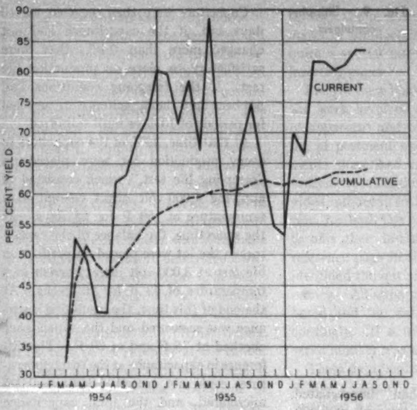


Fig. 7 (left). Resistor yield

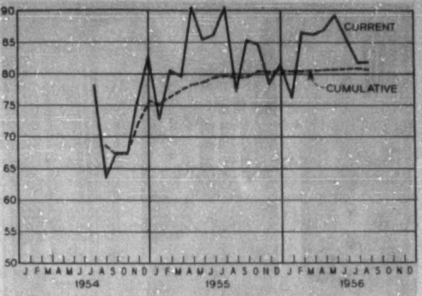


Fig. 8 (above). Inductor yield

Selection and Training of Inspection Personnel

The quality of a product naturally depends upon the skills, attitude, and integrity of the personnel making and inspecting it. It was realized that in order to develop the high degree of efficiency in the inspection organization necessary to ensure the integrity of the product, personnel of very high caliber would be required. The following qualifications were called for in these employees: 1. They would have to be experienced in similar or comparable work. 2. They would have to be precise, accurate and, above all, dependable. 3. In order to reduce the possibility of contamination and damage, they would have to be neat and careful. 4. They would require the ability to work in harmony with other employees, often as members of a team, in an environment where their work would be under constant scrutiny.

Most of the inspection employees selected to work at Hillside were transferred from the Kearny Plant and had an average Western Electric service of 12 years. They were hand-picked for the attributes outlined above, and the screening was performed by supervision through personal interviews supplemented by occupational tests given by the personnel department. These tests, which are in general use, are designed to evaluate background and physical characteristics, and they were given regardless of whether the employee had or had not previously taken them.

The following group of tests is an example of those given inspectors and testers of apparatus components:

1. Electrical: a-c—d-c theory and application.
2. Ortho-rater: Eye test for phoria, acuity, depth, and color.
3. Finger dexterity: Ability and ease of handling small parts.
4. Special: Legibility of handwriting, ability to transcribe data and to use algebraic formulas in data computations.

Inspection Plan

The general plan of visual and mechanical inspection consisted of:

1. Inspection of every operation performed, and in many cases partial operations, during the course of manufacture. This is of particular importance where the quality characteristics are hidden or inaccessible after completion of the operation.
2. Repeated inspection at subsequent points for omissions, damage, and contamination.
3. Rejection of product at any point where there was failure to obtain inspection or where the results of such inspection had not been recorded.

Most of the visual inspection was performed at the operators' positions to reduce to a minimum the amount of handling that could result in damage and contamination.

Visual inspection covered three general categories:

1. Inspection of work after some or all operations had been completed, such as the machining of parts.
2. Inspection at those points where successive operations would cover up the work already performed. An example of this is the hand winding of toroidal inductors where each layer of wire was examined under a microscope for such defects as twists, cracks, and crazes in enamel insulation, spacing and overlapping of turns, and contamination before the operator was allowed to proceed with another layer. While being inspected, the work remained in the holding fixture, which was hinged in such a manner as to permit inspection of both top and bottom of the coil. Inductors received an average of 13 and a maximum of 26 visual inspections during winding.
3. Continuous "over-the-shoulder" inspection, where strict adherence to a process was required or where it was impossible to determine, by subsequent inspection, whether or not specific operations had been performed. In these cases, the inspector checked the setup and facilities, observed to see that the manufacturing layouts were being followed, that the operations were being performed satisfactorily, and that specifications were being met.

Electrical Testing

The electrical testing, in itself, was not unusual for carrier apparatus, and it runs the gamut from d-c resistance through capacitance, inductance, and effective resistance, to transmission characteristics in the frequency band 20 to 174 kc. What was unusual were the extremely narrow limits imposed and the number and variety of tests involved as compared to those usually specified for commercial counterparts.

The following two examples will serve to illustrate the extreme measures taken to prove the integrity of the product:

The first example is one type of resistor wound with no. 46 mandrelated nichrome wire to a value of 100,000 ohms ±0.3%. This resistor received six checks for d-c resistance, five for instantaneous stability of resistance, and two for distributed capacitance, at various steps in the proc-

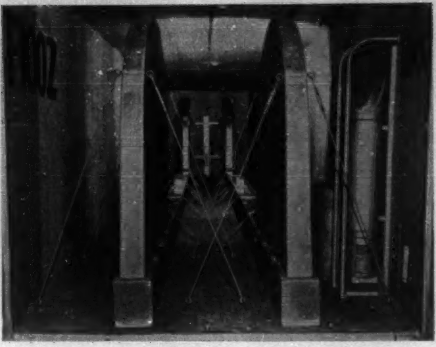


Fig. 9. Shipping containers

ess which included 6 days' temperature cycling for mechanical stabilization. This resistor was considered satisfactory, after final analysis of the test results, if: 1. the difference in any two of the six resistance readings did not exceed 0.25%, 2. the change in resistance during cycling was not greater than 0.02%, 3. the "instantaneous stability" (maximum change during 30 seconds) did not vary more than 0.01%. In addition, it was required that the distributed capacitance, minimum 7, maximum 10 micromicrofarads, should not differ from any other resistor by more than 2 micromicrofarads.

As the second example, for high-voltage paper capacitors, the 0.0004-inch-thick kraft paper, which constitutes the dielectric, was selected from the most promising mill lots which the manufacturers had to offer. This selection was based on the results obtained from tests that involve examination for porosity, conducting material, and conductivity of water extractions. These tests were followed by the winding and impregnation in Halowax of test capacitors. The test capacitors were then subjected to a direct voltage endurance test at 266 F for 24 hours.

Samples of prospective lots of paper, which have passed the foregoing test, were then used to wind another group of test capacitors that were subsequently impregnated with Aroclor and sealed; 1,500-volt direct current was then applied to the capacitors at 230 F for 500 hours. In case of failure, a second sampling was permitted.

After the foregoing tests had been passed, the supplier providing the particular mill lot was authorized to slit the paper. Upon receipt, six special capacitors were wound, using a group of six rolls of the paper being qualified. These capacitors were then impregnated, checked for dielectric strength at 3,000-volt direct current, and measured for capacitance and insulation resistance. The capacitors were then given an accelerated life test at 2,000-volt direct current, temperature 150 F, for 25 days. Each lot of six satisfactory test capacitors qualified six rolls of paper for use.

Product capacitors were then wound from approved paper, and the dry units checked for dielectric strength at 300-volt direct current. Capacitance was checked and units were then assembled into cans and ceramic covers soldered in place. Assemblies were pressurized with air, through a hole provided for the purpose, while the assembly was immersed in hot water to determine if leaks were present. Capacitors were then baked, vacuum dried, impregnated, pressurized with nitrogen, and sealed off. The completely sealed units were then placed in a vacuum chamber at a temperature of 150 F, 2-millimeter mercury, for 3 hours to check for oil leaks. Capacitance was rechecked and insulation resistance measured.

After 7 days, capacitors were unsealed to replenish the nitrogen that had been absorbed by the oil, resealed, and again vacuum leak tested. An X-ray examination was then made of each individual unit to verify internal mechanical conditions. Capacitors were then placed in a temperature chamber and given the following cyclic treatment:

$$\left.\begin{array}{l}\text{16 hours at 150 F}\\ \text{8 hours at 75 F}\\ \text{16 hours at 0 F}\\ \text{8 hours at 75 F}\end{array}\right\}\text{1 cycle}$$

At the end of 10 days, or 5 cycles, the insulation resistance and conductance were measured and a norm established for capacitance.

Capacitors were then recycled for 10 days and, if the capacitance had not changed more than 0.1%, they were satisfactory to place on production life test. If the foregoing conditions had not been met, the capacitors were recycled for periods of 10 days until stabilized.

At that time, 10% of the capacitors in every production lot were placed on "sampling life test," which consisted of applying 4,000-volt direct current in a temperature of 150 F for 25 days. At the same time, the balance of the capacitors in the lot were placed on production life test at 3,000-volt direct current in a temperature of 42 F for 26 weeks. At the end of this time, the insulation resistance was measured and the capacitance checked at 75 F and at 39 F. The difference in capacitance at the two temperatures could not exceed +0.001, −0.005 microfarad, and the total capacitance could not exceed maximum 0.3726, minimum 0.3674 microfarad. The capacitance from start to finish of the life test could not have changed more than ±0.1%.

If all of these requirements were satisfied, the particular lot of capacitors described was considered satisfactory for use in the product. The foregoing examples are typical of the procedures evolved for ensuring, to the greatest degree possible, the long, trouble-free life of all apparatus used in the repeater.

RADIOISOTOPE TEST

There were many new and involved tests developed and applied to the manufacture of repeaters. One of the unique tests is the use of a radioisotope for detecting leaks under hydraulic pressure.

The initial closure operations consisted of brazing into each end of the repeater housing a Kovar-to-glass seal. These seals are equipped with small-diameter nickel tubulations which were used to flush and pressurize the repeaters with nitrogen. After these operations had been performed, one of the tubulations was pinchwelded, overbrazed and coiled down into the seal cavity. The repeater was then placed in a pressure cylinder with the open tubulation extending through and sealed to the test cylinder. A mass spectrometer was then attached to the tubulation and the test cylinder pressurized with helium at 10,000 psi. At the conclusion of this test the repeater was removed from the test cylinder and, after breaking the desiccator diaphragm, the remaining open tubulation was pinchwelded and overbrazed. At this point, it became necessary to determine whether the final pinchweld and overbrazing would leak under pressure.

Since there was no longer any means of access to the inside of the repeater, all testing had to be done from the outside. This was accomplished by filling the glass seal with a solution of radioisotope cesium 134, which was retained by a fixture. The repeater was then placed in a test cylinder and hydraulic pressure applied, which was transmitted to the radioisotope in the fixture. After 60 hours under pressure, the repeater was removed from the cylinder and the seal drained and washed. An examination was then made with a Geiger counter to determine if any of the isotope had entered the final weld.

The washing procedure, after application of the isotope solution, involved some 60 operations with precise timing. In the case of the repeater at the rubber seal stage where both ends were tested, it was desirable that these operations be performed concurrently. This was accomplished by recording the entire process on magnetic tape which, when played back, furnished detailed instructions and exact timing.

Raw-Material Inspection

As might be expected, raw materials used in the project were very carefully examined and nothing left to chance. Every individual bar, rod, sheet, tube, bottle, or can of materials was given a serial number and a sample taken from each and similarly identified. Each sample was then given a complete chemical and physical analysis before each corresponding piece of material was certified and released for processing. In many cases, the cost of inspection far exceeded the cost of the material. However, the discrepancies revealed and the assurance provided more than justified the expense.

Detailed records of all raw-material inspection were compiled and furnished to the responsible raw-material engineer who examined them, critically, as an additional precaution before the material was released to the shop.

Inspection Records

To eliminate, as much as possible, the human element in providing assurance that all prescribed operations had been performed satisfactorily and inspected properly, and the results recorded, means were established to compile a complete history of the product concurrent with manufacture. This was accomplished through the provision of permanent data books of semilooseleaf design, which require a special machine for removing or inserting pages.

Each of these books covered a portion of the work involved in producing a piece of apparatus and contained a sequential list of pertinent operations and requirements prescribed in the manufacturing process specifications. Space was provided, adjacent to the recorded information, for both the operator and inspector to affix their initials and the data. A reference page in the front of each book identified the initials with the employees' names. All apparatus was serially numbered and the data were identified accordingly. If a unit was rejected, that serial number was not reused.

These data books, in addition to establishing a complete record of manufacture, provided a definite psychological advantage, in that people were naturally more attentive to their work when required to sign for responsibility.

Quality Accuracy

As pointed out previously, every precaution was exercised in selecting and training inspection personnel assigned to the project. However, it was realized at the outset that human beings are not infallible and that insurance, to the greatest degree possible, would have to be provided against the probability of errors in observation and judgment. Quality accuracy evaluation procedures were therefore established for determining the accuracy of each inspector's performance.

Quality accuracy checking was performed by a staff of five inspection representatives and involved an examination of the work performed by inspectors to determine how accurately it was inspected. Materials which the inspector accepted and those which had been rejected were both examined.

Verification and Summary of Data

As an added measure of assurance as to the integrity of the product, procedures were established for verifying and summarizing the inspection records for each serially numbered component, up to and including complete repeaters.

Verification involved a complete audit of the inspection records to provide assurance that all process operations were recorded as having been performed satisfactorily, that the prescribed inspections had been made, and the the recorded results indicated that the product met all of the specified requirements. This work was performed by a group of six inspection representatives who had had considerable experience in all phases of inspection and inspection records.

As the verification of a particular piece of apparatus proceeded, a verification report was prepared which, when completed, contained the most pertinent inspection data, such as:

1. Recorded measurements of electrical parameters.
2. Values calculated from measurements to determine conformance.
3. Confirmation that all process and inspection operations had been verified.
4. Identification (code numbers and serial or lot numbers) of materials and components entering into the product at each stage of manufacture.

The verification report usually listed the data for 20 serial numbers of a particular code of apparatus, along with the specified requirements. Included, also, was a cross reference to all the inspection data books involved so that the original data could be located easily. These verification reports were prepared for all apparatus up to and including the finally assembled and tested repeaters.

The following gives an indication of the number of items examined in the verification of one complete repeater:

Items verified in data books	17,593
Items verified on recorder charts	1,142
Calculations verified	1,580
Total	20,315
Number of entries on verification reports	4,070

Verification reports, in addition to presenting the pertinent recorded data, provided a "field" of 20 sets of measurements from which it was easily possible to spot a questionable variation. For example, it was the adopted practice on this project to examine critically any characteristic of a piece of apparatus, in a universe of 20, which varied considerably from the rest, despite the fact that it was still within limits.

While the number of cases turned up in the verification process which resulted in rejection of product are relatively few, it is believed that the added insurance provided, and the psychological value obtained, considerably outweighed the cost.

Power-Feed Equipment for the North Atlantic Link

G. W. MESZAROS H. H. SPENCER

Synopsis. Precise regulation of the direct current which provides power for the undersea repeaters in the new transatlantic telephone cable is necessary to maintain proper transmission levels and to assure maximum repeater tube life. The highest possible degree of protection is needed against excessive currents and voltages under a wide variety of possible fault conditions. Furthermore, to minimize the dielectric stresses, a double-ended series-aiding power feed must be used, and the balance of these applied voltages must be maintained despite substantial earth potentials. This paper describes the design features which were employed to attain these objectives simultaneously, while eliminating, for all practical purposes, any possibility of even a brief system outage because of power failure.

THE PRINCIPAL objectives in the power plant design for the transatlantic cable system were as follows:

1. To stress reliability in order to guarantee continuous d-c power to the electron tubes that form an integral part of the submerged repeaters. This is essential, not only to be able to maintain continuous service, but also to prevent cooling and contraction of the repeater components, especially the tubes.

2. To provide close d-c cable-current control to ensure constant cathode temperature and regulated plate and screen potentials for the repeater tubes. These operating conditions are essential to obtain both maximum life from these tubes and constant transmission level.

3. To control and limit the applied d-c cable potentials to minimize the dielectric stresses. The life of certain capacitors in the repeaters is critically dependent upon these stresses. Moreover, momentary high potentials increase the chances of corona formation and insulation breakdown.

4. To protect the cable repeaters from the excessive potentials or currents to which they might be subjected after an accidental open or short circuit in the cable.

5. To compensate for earth potentials up to 1,000 v (volts), of either polarity, which may develop between the grounds at Oban and Clarenville during the magnetic storms which accompany the appearance of sun spots and the aurora borealis.

6. To provide adequate alarms and automatic safety features to ensure safe current and voltage conditions for both the cable and the operating personnel.

Design Requirements

Reliable Cable Power

The first basic problem of design was to select a reliable source of d-c power for energizing the cable repeaters. Although a string of batteries, on continuous charge, is perhaps the most dependable source of direct current, such an arrangement is not attractive here. A complex set of high-potential switches would be required for removing sections of batteries for maintenance and replacement purposes. Protection of the repeater tubes from damage during a cable short circuit would be difficult. Facilities to accommodate changing earth potentials would be cumbersome. Furthermore, the problem of hazards to personnel would be serious.

The use of commercial a-c power with transformers and rectifiers to convert to high-potential direct current would expose the cable to power interruptions even with a stand-by diesel-driven alternator, because of the time required to start the engine. A diesel plant could be operated on a continuous basis, but this prime power source would also present a considerable failure hazard even with the best maintenance care. The 2-motor alternator set, used so successfully in the Bell System's type *L* carrier telephone system, was adopted as representing the most reliable continuous power source available. This set normally operates on commercial a-c power, but when this fails, the directly coupled battery-operated d-c motor quickly and automatically takes over the drive from the induction motor to prevent interruption of the alternator output. Here the storage battery is still the foundation for continuity, but at a more reasonable voltage.

As described later, the possibility of a system outage resulting from failure of this 2-motor alternator set has been essentially eliminated by using two of these sets, cross-connected to the rectifiers which supply power to the two cables, with a continuously operating spare for each set, automatically switched in upon failure of the regular set.

The regulating features of the rectifiers will be described in a later section. In the present discussion of reliability it is sufficient to note that series regulating tubes are used, which are capable of acting as high-speed switches, through which two rectifiers can be paralleled. Thus, either rectifier can accept instantaneously the entire load presented by the cable. In each regulator, the series tubes carrying the cable current are furnished in duplicate and connected in parallel to share the cable load, a single tube being capable of carrying the entire load. These current regulators are operated from separate a-c sources to protect against loss of cable power because of failure of one of the sources of a-c power.

Cable Potentials

To minimize the cable potentials, half of the d-c power is supplied at each end of each cable, the supplies being connected in series aiding. With this arrangement, as shown in Fig. 1, the d-c cable potential at one end of each cable is positive with respect to ground, while at the other end the potential is negative. This places the maximum potential and risk on the repeaters near the shore ends, which are more readily retrieved, while the repeaters in the middle of the cable, in deeper water, have potentials very near to ground. The power equipment would be simpler with a single-ended arrangement, but at the penalty of doubling the dielectric stresses in the entire system, which would be prohibitive. A balanced power feed could have been attained at the expense of power-separation filters in the middle of the cable, or a shunt impedance of appropriate size at the mid-point. The resulting complications, including difficulty in the location of a cable fault, could not be justified for the sake of simplification of power-plant design and operation.

The requirement that minimum cable potentials be maintained during and after severe earth potential disturbances necessitates variable output voltages from the supplies at both ends, and this introduces problems in continuous voltage balance and regulation stability. The design features which yield the required performance are described in a later section.

D-C Cable-Current Regulation

The salient requirements in performance of the constant current regulator are listed as follows:

1. The regulator must have extremely fast response to hold the cable current within a few milliamperes of its nominal value should a short circuit develop in

Vol.
Pt. 1
1956

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the Winter General Meeting, New York, N. Y., January 21–25, 1957.

G. W. Meszaros and H. H. Spencer are with the Bell Telephone Laboratories, New York, N. Y.

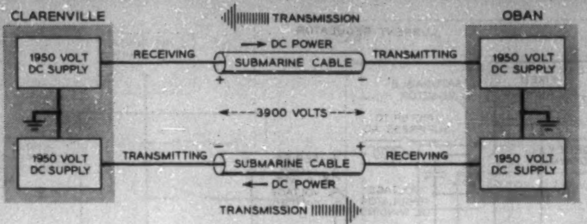


Fig. 1. Cable voltage supply

the cable. Thus, damage to the heaters of the repeater tubes, as well as excessive induced transient voltages in the repeater transformers, are avoided. The probability of a short circuit is greater near the shore ends where the water is shallow and sea traffic is a factor. The regulator must be capable of absorbing the reduction in power to the cable, while maintaining current control under normal conditions. This sudden exchange in power from cable to regulator may be as much as 2,000 v at 0.25 amp (ampere).

2. The cable current should be maintained constant within 0.2% (per cent) of its nominal value for normal variations in a-c supply, gradual earth-potential changes, and ambient-temperature changes. This degree of regulation allows an adequate safety factor in maintaining a constant transmission level.[1]

3. The regulators, with the power-separation filters and the rectifier filters, must limit the power-supply noise at the cable terminals to a peak-to-peak value less than 0.02% of the d-c supply potential.

4. The cable current must be adjustable over a range of 225 to 245 ma (milliamperes) to compensate for repeater-tube aging.[2]

5. The regulators must operate in parallel in such a way as to ensure continuity of power should one fail or be removed from service for maintenance. This of course implies that regulators can be switched in and out of service without causing surges in the cable current or voltage.

6. The series-aiding arrangement, with rectifiers at each end of the same cable, must be stable.

7. The regulators should be capable of being serviced at low potentials when in the test position to protect maintenance personnel.

8. The current regulator should be of the fail-safe type so that impairment of any of the regulator components will not permit excessive rise in cable current. In the event of component trouble, an aural or visual alarm should occur.

It was decided that a high-speed electronic constant-current regulator backed up by a slower speed servo system, as shown in Fig. 2 and discussed in detail later, would best meet these requirements. In this way, fast response with high gain is combined with wide regulating range, yet the efficiency is high and the load-handling capacities of the various components are held to a minimum. With regard to simpler alternatives, the electromechanical-type current regulator which uses relays and a motor-driven rheostat is too slow to protect the repeater tubes from a cable short circuit and its accuracy is insufficient to meet the regulation requirements. The all-magnetic type of regulator is possibly most dependable, but it does not readily provide either the speed of response or the wide regulating range needed.

General Description

Prime and Stand-by Power Source

Commercial service is considered the normal prime source of power for the cable, although at the Clarenville terminal commercial power was not available at the time of installation. Anticipating this condition, a reserve plant consisting of three 60-kw diesel alternators was installed and the distribution circuits were arranged, as shown in Fig. 3, to provide partial or total use of the commercial service. Initially, all cable power was supplied by diesel operation, alternating the prime movers on a weekly basis. These sets are paralleled manually, when they are interchanged, to prevent an interruption in the 60-cycle sypply. It may be noted that engine no. 1 is arranged as an automatic stand-by whether prime power is provided by diesels or by commercial service.

The switching and distribution arrangements are designed to be essentially failure-proof. At Clarenville, for example, two a-c distribution cabinets, each capable of being fed from two sources, were provided in separate locations. The normal source through engine no. 1 control bay can be readily by-passed directly to the manual diesels should engine no. 1 control bay be disabled. Furthermore, allocation of charging rectifiers, control circuits, a-c motors for continuity sets, etc., has been made in such a manner that loss of one cabinet alone will have minimum effect on the cable power supplies or office loads. At Oban, where 50-cycle commercial service is normally used, special distribution arrangements have been provided to give maximum power-supply reliability, with three manually operated 50-cycle 90-kw diesel-alternator sets arranged for stand-by service. The diesels, at this terminal, are larger to care for greater local power loads.

Continuous A-C Power From 2-Motor Alternators

At both cable terminals, two reliable a-c busses supply power to the d-c cable regulating bays. Each of these busses is fed from a continuously operated, self-excited, single-phase 230-v alternator, normally driven by a 3-phase induction motor on the same shaft with a 130-v d-c motor. Each regular alternator is backed up by a similar emergency alternator running at no load. A fifth motor alternator is provided, which can be used whenever any other set is out of service for routine maintenance or repair.

As alternator loads are essentially constant and since induction motor speeds are fairly insensitive to power-supply voltage variations, alternator outputs are set by fixed adjustments of their field rheostats. Supply voltages are monitored to control automatic transfer to d-c motor drive whenever the supply voltage drops below 80% of the normal value. Fig. 4 shows the normal running circuit for an alternator set, with the d-c motor connected to the battery through a resistance of 75 ohms inserted in the armature circuit. The field resistance *FR* is preset so that, when the battery is driving the set, the speed matches that of the a-c drive when the battery voltage is at the mean discharge value. During a-c drive, the emf (electromotive force) generated in the d-c motor armature is a few volts below that of the battery. Accordingly, when the fast-acting contactor *AR* shorts the 75-ohm resistance in the armature circuit, the motor finds itself essentially at the desired operating flux condition, and a smooth pickup of drive occurs. Oscillograms indicate that the interval between failure of a-c power and operation of contactor *AR* is less than 0.1 second. During the transfer from a-c to d-c drive, the change in the nominally 230-v output is less than 5 v.

Return to a-c drive is delayed approximately 20 seconds after the a-c supply voltage has returned to normal to allow time for the alternating current to stabilize. Fixed field settings for both d-c

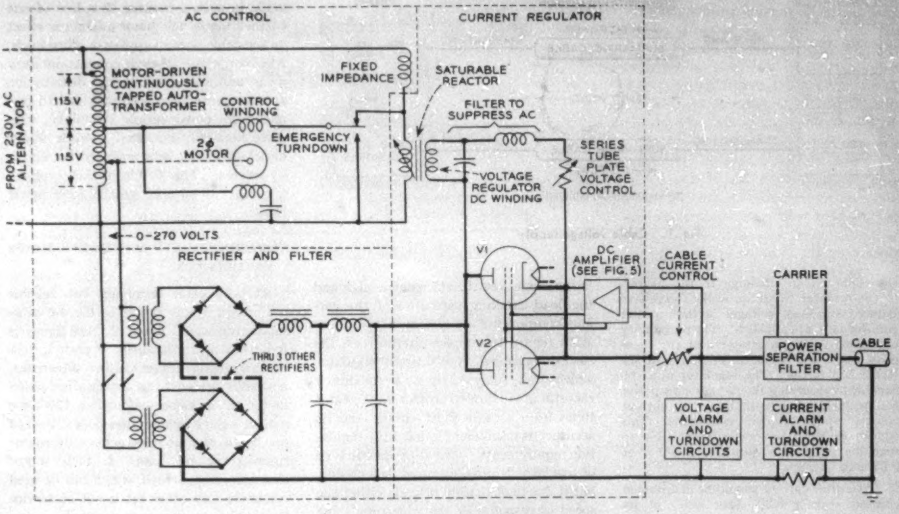

AC CONTROL
CURRENT REGULATOR
FROM 230V AC ALTERNATOR
115 V
115 V
MOTOR-DRIVEN CONTINUOUSLY TAPPED AUTO-TRANSFORMER
CONTROL WINDING
2φ MOTOR
EMERGENCY TURNDOWN
FIXED IMPEDANCE
SATURABLE REACTOR
FILTER TO SUPPRESS AC
SERIES TUBE PLATE VOLTAGE CONTROL
VOLTAGE REGULATOR DC WINDING
0-270 VOLTS
RECTIFIER AND FILTER
THRU 3 OTHER RECTIFIERS
V1
V2
DC AMPLIFIER (SEE FIG. 5)
CABLE CURRENT CONTROL
CARRIER
POWER SEPARATION FILTER
CABLE
VOLTAGE ALARM AND TURNDOWN CIRCUITS
CURRENT ALARM AND TURNDOWN CIRCUITS

Fig. 2. Simplified circuit of the regulating system

motor and alternator fields provide simple control arrangements without the overspeed or overvoltage hazards which automatic regulators might add. Alternator output is monitored, however, to give alarms for voltage changes exceeding ±5% and to control transfer to d-c drive if the output should drop more than 10% for any reason. When the latter occurs, the machine locks on d-c drive. This feature guards against a-c motor failure or low a-c drive speed because of low supply frequency without low supply voltage.

Failure of the alternator output after transfer to d-c drive causes the set to stop and automatically transfer the load to its emergency alternator. This transfer causes a break in the alternator supply to its bus, but cable power is maintained constant by the parallel d-c regulating bay fed from the other alternator bus.

Battery Plants and Distribution

Battery power for d-c motor drive is supplied from a 66-cell 1,680-amp-hr (ampere-hour) battery at Clarenville, and from two 68-cell 1,680-amp-hr batteries at Oban. The latter station has double capacity to provide stand-by power for the a-c supplies to the inland transmission equipment.

At both cable terminals, control battery for the small alternator plants and the cable d-c regulating equipment provides 24 v and is split so that a fuse or a battery failure on either supply will not interrupt cable power. To guard against so remote a hazard as loss of a common battery for this vital control, two separate 24-v power plants have been provided with one half of the critical control circuits furnished from each plant.

In addition to supply d-c motor power, the 130-v battery at Clarenville supplies current to the carrier terminal and test equipment through voltage-dropping resistors which are normally in the circuit to hold the load voltages below a maximum of 135 v. These resistors are automatically short-circuited to maintain a minimum load voltage of 125 v when the battery voltage drops. A voltage-detecting relay also steps the fixed field adjustment of each d-c motor when the battery nears its final discharge voltage so that d-c motor speed is kept within about ±3% of normal a-c drive speed during battery operation. The 66-cell battery is floated at 143 v by means of voltage-regulated rectifiers and, after a discharge, is recharged by automatic operation of a regulated 100-amp motor generator set. As indicated in Fig. 3, the rectifiers for this plant are connected so that loss of one service cabinet will still leave sufficient charging capacity to float the load from the other cabinet.

Rectifiers and Associated Controls

As mentioned previously, each cable is supplied at all times by two regulating bays in parallel, each operating from a separate a-c source and each capable of taking over the cable load should the other fail. Thus, failure of an alternator supply or of a regulating bay itself will not interrupt the cable power. A spare regulating bay for each cable is arranged to replace either of the two regular bays and may be connected in parallel with the other two without overloading the associated 2.5-kva alternator.

As shown schematically in Fig. 2, the a-c supply to the rectifiers is controlled by a continuously tapped variable autotransformer, operated normally by a 2-phase low-inertia reversible motor, with provision for manual adjustments also. The autotransformer output is stepped up and rectified by a series arrangement of rectifiers, each to give a maximum of 550 v each, for a total of 2,750 v when the autotransformer is at the upper limit. The high-voltage rectifier output is filtered and supplied to the cable through the current regulating unit, cable connecting switch, and common cable control circuit. The cable current is regulated by controlling the plate to cathode drop across the electron tubes through which this current flows. A d-c amplifier, which derives its signal from a resistance in series with the

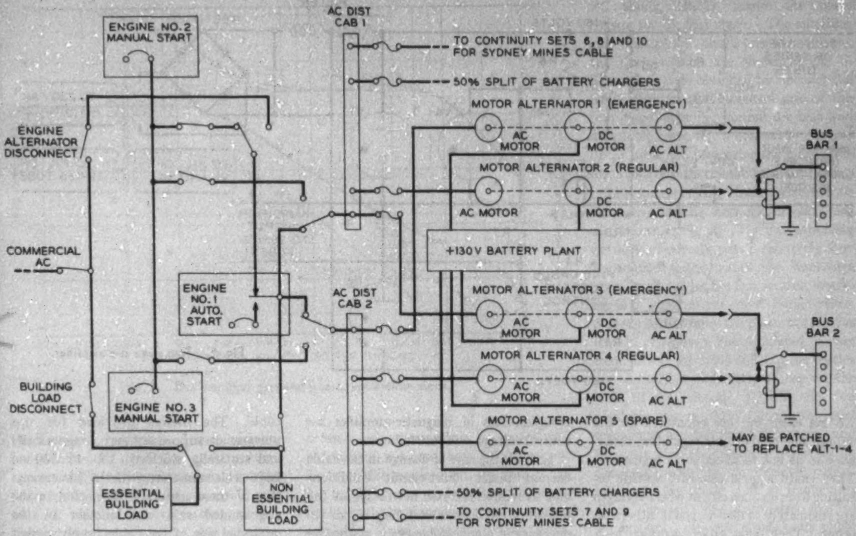

ENGINE NO. 2 MANUAL START
AC DIST CAB 1
TO CONTINUITY SETS 6, 8 AND 10 FOR SYDNEY MINES CABLE
50% SPLIT OF BATTERY CHARGERS
MOTOR ALTERNATOR 1 (EMERGENCY)
ENGINE ALTERNATOR DISCONNECT
AC MOTOR
DC MOTOR
AC ALT
BUS BAR 1
MOTOR ALTERNATOR 2 (REGULAR)
AC MOTOR
DC MOTOR
AC ALT
+130V BATTERY PLANT
COMMERCIAL AC
ENGINE NO. 1 AUTO. START
AC DIST CAB 2
MOTOR ALTERNATOR 3 (EMERGENCY)
AC MOTOR
DC MOTOR
AC ALT
BUS BAR 2
MOTOR ALTERNATOR 4 (REGULAR)
AC MOTOR
DC MOTOR
AC ALT
BUILDING LOAD DISCONNECT
ENGINE NO. 3 MANUAL START
MOTOR ALTERNATOR 5 (SPARE)
AC MOTOR
DC MOTOR
AC ALT
MAY BE PATCHED TO REPLACE ALT-1-4
ESSENTIAL BUILDING LOAD
NON ESSENTIAL BUILDING LOAD
50% SPLIT OF BATTERY CHARGERS
TO CONTINUITY SETS 7 AND 9 FOR SYDNEY MINES CABLE

Fig. 3. Continuous a-c power

total cable current, varies the grid bias of the series regulating tubes. The series tubes have a control-drop range from about 150 v to the full supply voltage, the d-c amplifier being capable of driving the regulating tubes of cutoff. However, to protect the series tubes and to keep them operating at practically a constant plate voltage of 300 v, a servo system automatically raises and lowers the output from the autotransformer by means of the motor-driven a-c control unit, whenever the series-tube plate voltage varies more than 25 v from the normal value. For example, a 2% change in input alternating current would change the rectifier output of 2,300 v, which corresponds to a cable supply voltage of 2,000 v, by 46 v, which would increase the series-tube drop to 346 v. This increase in voltage would raise the signal current through the control winding of a saturable reactor which forms one leg of a balanced bridge in the a-c motor control circuit, and thus cause the autotransformer to be driven down, lowering the rectifier output until the series-tube drop is restored again to approximately 300 v.

OVERVOLTAGE AND OVERCURRENT PROTECTION AND ALARMS

While the power for the cable is electronically regulated, protective features are provided to guard against abnormal cable current or voltage. The first order of protection is a ±2% cable-current alarm, given by a voltage relay, which operates from the drop across a resistor in the ground-return circuit to the d-c bays. A second voltage relay, set for 5% high cable current, also monitors the current in the ground-return side, and with the current monitors in the high-voltage side it limits the cable current by operating the motor-driven autotransformers until the current is within 5% of normal. The voltage of the ungrounded side, however, is much too high for directly connected voltage or current relays. Therefore, magnetic amplifiers have been used to obtain isolated metering of the cable current. Two of these devices measure the current in the ungrounded side of the common power-supply lead to the cables. Of the three current monitors available, two must operate before turndown functions to prevent false turndown because of faulty metering.

Voltage protection is provided by means of magnetic amplifiers in shunt across the common power supply to the cable. Here, three monitors are arranged so that any two can reduce power, by means of the turndown control, if the voltage rises to the maximum allowable value for which the voltage relays are adjusted. These monitors draw about

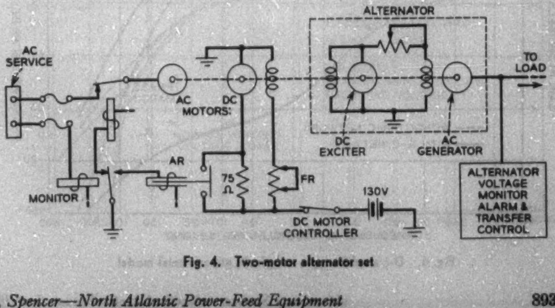

ALTERNATOR
AC SERVICE
AC
DC
MOTORS:
TO LOAD
DC EXCITER
AC GENERATOR
AR
75 Ω
FR
MONITOR
130V
DC MOTOR CONTROLLER
ALTERNATOR VOLTAGE MONITOR ALARM & TRANSFER CONTROL

Fig. 4. Two-motor alternator set

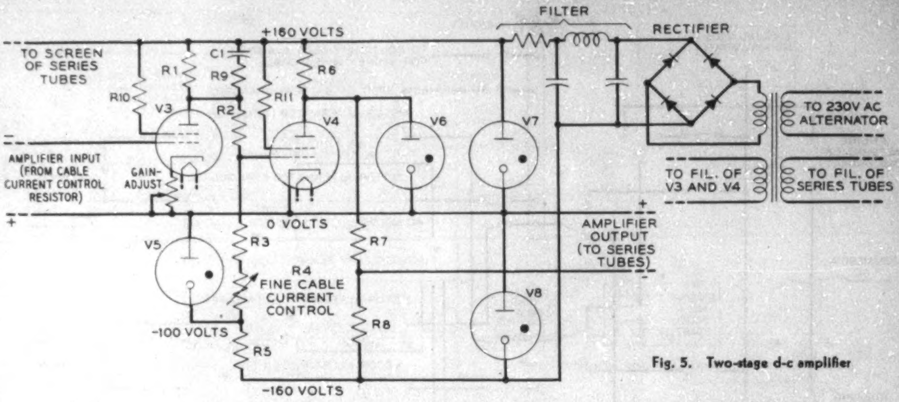


Fig. 5. Two-stage d-c amplifier

1.5 ma each, but are connected on the supply side of the cable-regulating resistance so as not to affect cable regulation. They guard against excessive voltage resulting from an open circuit where voltages approximately 4,000 v could otherwise occur. They also guard against high voltage caused by earth potentials or unbalance between voltages at opposite ends of the cable. When set for a ceiling voltage of 2,600 v, a maximum of 3,000 v occurs on open circuit on the first rise, after which the voltage holds within 2,600 ±200 v as the turndown relays operate and release to maintain the ceiling voltage. A fourth magnetic-amplifier voltage-detecting relay provides an alarm for ±5% excursions in cable voltage from the normal value. Other alarms are provided to indicate low output in either of the two parallel regulating bays, relay troubles, loss of magnetic-amplifier a-c control voltage, and fuse failures.

To limit the rate of change in the cable current under short-circuit conditions and to reduce the rise in voltage at the repeaters on open-circuit failure, an inductance of about 36 henrys is connected in series with the cable circuit, and physically close to the cable termination, so that any failure in the power supply would have the advantage of this surge-limiting element.

METERING

Metering of the cable current is a very important part of the power-plant design. Not only are the cable-current ammeters needed to set the value of current desired, but also their ability to indicate absolute current values assists in obtaining stable regulation between the two ends of the cable. The meters provided for this purpose are suppressed-zero, magnetically and statically shielded, 150- to 300-ma large-scale ammeters, with 0.5% accuracy. One of these meters is connected in the ungrounded side and another in the grounded side of the cable supply circuit to provide an accuracy check and to indicate any ground-leakage current in the supply circuit. They are connected to highly accurate, 1.0-ohm 4-terminal shielded resistors, acting as shunts in the cable-current circuits, with their shunt leads arranged for switching to a calibration box for checking accuracy and adjustment. This box uses a Weston laboratory standard cell, which is essentially a single-point potentiometer with the usual galvanometer, acting as a calibration standard at the 225-ma point. Meters calibrated at Oban for 225 ma were expected to be within 0.2% or 0.5 ma of those similarly calibrated at Clarenville, and, at present, are within about 0.2 ma.

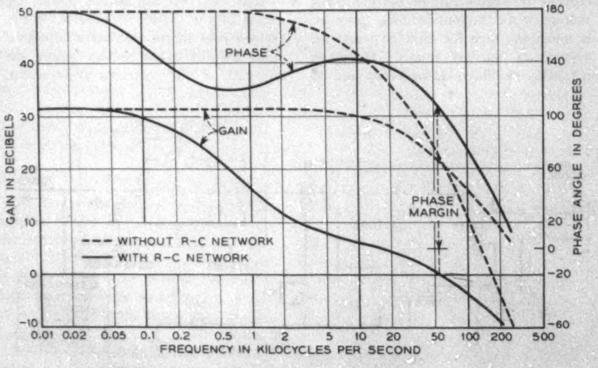


Fig. 6. D-c amplifier gain and phase, experimental model

The cable current is also indicated by a recording ammeter. Meters in each regulating bay indicate the division of current between paralleled regulators. These meters have only 1.0 % accuracy but are satisfactory for adjusting load balance between parallel bays and also are used in the turn up of power on a particular bay.

Cable voltage is read on a large-scale voltmeter reading 0 to 3,000 v and having ±0.25% accuracy. Since the accuracy is not critical, this meter is not arranged for calibration. However, the series resistors incorporate 6,000-v components for an extra degree of reliability. The voltmeter, with its series resistance, is normally connected across the common cable power supply, ahead of the cable

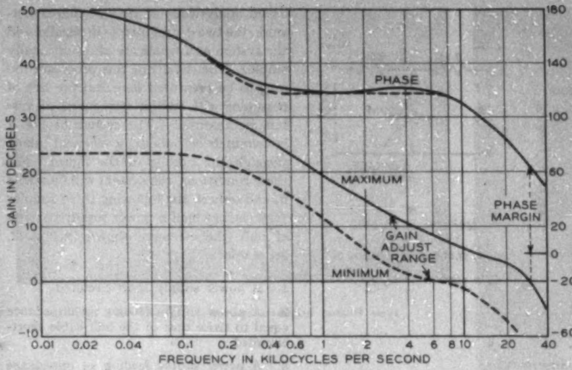


Fig. 7. D-c amplifier gain and phase, production model

current regulating point. It can, however be switched to read the cable voltage nearer the cable termination. In the latter position, it reduces the cable current by about 1.0 ma, causing an unbalance in cable regulation, and therefore is not normally left in this position. The cable supply voltage is also indicated by a recording voltmeter.

Other meters are provided to indicate series-tube plate voltages, d-c rectifier voltages, a-c input voltages, series-tube currents, test currents for adjusting magnetic-amplifier operating limits, and the difference in current between the positive and negative power supplies to ground. Since many of these instruments operate at high potential a special design was used, with the operating mechanism and scale depressed in the instrument case approximately 1.0 inch. This eliminated possible electrostatic effects on the instrument pointers, and served as a safety measure.

D-C Regulation

D-C Amplifier

The direct-coupled 2-stage amplifier, shown in Fig. 5, is characterized by potentiometer coupling and cold-cathode gas-tube voltage stabilizers. The biases are selected high enough so that linear operation is assured even for a short circuit at the cable terminal. As is apt to be the case in a direct-coupled amplifier, cathode temperature in the low-level stage is critical. In this amplifier, a 5% change in heater voltage results in a 0.5-ma change in the cable current.

The required precision of current regulation in these power supplies can be expressed as representing a source impedance of not less than 100,000 ohms. To meet this requirement, the gain of the d-c amplifier was made as large as practicable with the plate and screen potentials available from gas-tube regulators. Interstage network impedances are high to reduce the shunt losses, and are proportioned to provide the large biases mentioned previously. Gain adjustment is provided by a variable resistance in series with the cathode of the first stage.

Shaping of the loop gain and phase characteristics to obtain margins for stable operation is accomplished by means of the *RC* shunt, (R_9C_1), across the plate resistor for the first stage. The amplifier gain and phase characteristics without this compensation are shown in Fig. 6. These data were obtained by opening the feedback loop at the control grid of the first stage, applying normal d-c bias and a variable frequency a-c signal to the grid of the tube, and measuring the magnitude and relative phase of the return signal.

The corresponding characteristics with the compensating network in place are also shown in Fig. 6. The compensating network effectively puts a relatively low-impedance shunt across the interstage network at the higher frequencies, resulting in a "step" in the gain characteristic. A secondary effect is the phase shift in the transition region. The calculated "corner frequencies" are 2,800 and 195 cps (cycles per second), respectively, chosen on the basis of the criteria: 1. little effect on regulator gain at 100 or 120 cps, the most prominent rectifier ripple frequency, and 2. a gain step of something about 20 db (decibels) with no appreciable contribution to the phase shift at frequencies above 30 kc. The calculated loss at 120 cps is 1.2 db, with a maximum phase shift of about 60 degrees at the median frequency. These results agree quite well with the measured data plotted in Fig. 6.

As indicated in Fig. 6, the phase margin at the gain crossover frequency of 55 kc was somewhat over 100 degrees for the experimental model on which these measurements were made. The gain margin could not be measured readily, but is

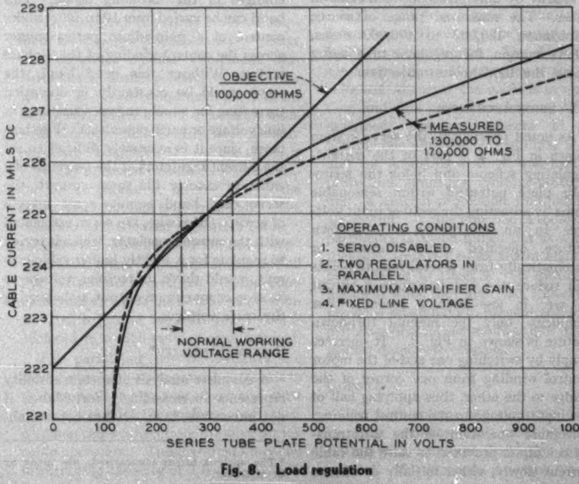


Fig. 8. Load regulation

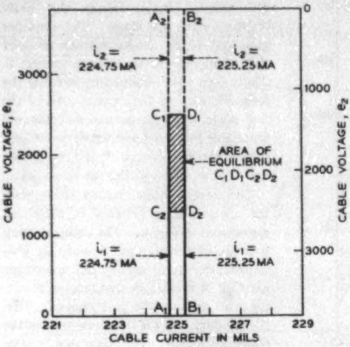
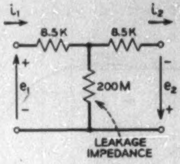


Fig. 9. Equilibrium diagram



MAPPING EQUATIONS
(VOLTS & MA)

$$e_2 = 17.00036\,i_1 - 1.00004\,e_1$$

$$i_2 = 1.00004\,i_1 - e_1/200{,}000$$

clearly substantial. On production units, larger wire sizes and longer lead lengths resulted in lesser, but still satisfactory stability margins, as shown in Fig. 7, where the phase margin is somewhat over 60 degrees. Fig. 7 also shows the characteristics at the extremes of gain control, where the range of control is about 8 db.

Fig. 8 shows the measured performance of the d-c regulators, the servo system being disabled to obtain a plot of the performance of the d-c amplifier and associated circuits. Twenty-two regulator units were manufactured and measured, and the curves of Fig. 8 show the extreme limits observed, with the differences between individual regulators caused primarily by differences between electron tubes. The measured range of source impedance, 130,000 to 170,000 ohms, allows margin for regulator-tube aging above the 100,000-ohm objective.

A-C SERVOMECHANISM

As noted previously, the servo system shown in Fig. 2 is part of the current-regulating scheme and holds the series-tube plate potential within reasonable limits by adjusting the rectifier input voltage. In an emergency, a turndown feature, operated either manually or automatically from several remote points will reduce the autotransformer output to zero in less than 2 seconds. For simplicity, only the manual turndown feature is shown in Fig. 2. It operates simply by switching one end of the motor control winding from one corner of the bridge to the other, thus applying half of the input voltage to the control winding.

Manual operation of the autotransformer tap is provided to raise the cable current slowly, either initially or after a turndown. In manual operation, a dynamic brake which consists of a short circuit on the motor control winding prevents the motor from creeping or coasting when the operator releases the handwheel, as it otherwise would since the fixed phase of the 2-phase motor is always energized. The turndown feature takes precedence over the short circuit of the motor control winding, automatically, to energize the motor should the operator inadvertently cause abnormally high cable voltage or current.

One essential feature of the servo design is the dead band of the series-tube plate voltage in which the servo remains stationary, even though there are small changes in the incoming signal. This band can be varied from 10 to 100 v under control of a gain-adjust potentiometer across the control winding of the 2-phase motor. Without this dead band, the servo would be constantly in operation correcting for small random variations in line voltage or earth potentials. Furthermore, since it is extremely difficult to set the current regulators at the two ends of a cable to exactly the same current, the servo dead band permits some margin of error. Otherwise, the servo associated with the current regulator, which is trying to regulate for a slightly higher cable current, would drive its rectifier voltage to its stop or maximum output, unbalancing the cable voltages.

SYSTEM STABILITY*

A complete analysis of system stability represents an exceedingly formidable, if not impossible, task. It has been established analytically that, for a linear network, the two d-c regulators in parallel and the system as a whole are unconditionally stable. The details of this proof are too long to be presented here, but the line of reasoning with respect to the over-all system is presented. The system of Fig. 1 is symmetrical about a vertical plane through the middle of the figure. For these conditions, the system will be stable if, and only if, the following three simpler systems, in which a power supply consists of only the elements shown in Fig. 2, are stable:

1. A power supply short-circuited.
2. A power supply feeding an impedance equal to twice that of the half cable short-circuited.
3. A power supply feeding an impedance equal to twice that of the half cable open-circuited.

The transfer function of the servomechanism was measured over the frequency range of principal interest, zero to one cps. The behavior near zero frequency was determined from the asymptotic slope of the unit step response. During time-domain measurements, it was quite apparent that the a-c control loop could be considered as a linear network only in an approximate sense, and thus that the analytical results were primarily useful in interpretation of observed behavior of the system. In this frequency range the d-c amplifier gain is a real constant, flat gain, and negligible delay as previously shown. Therefore, only the a-c servo-feedback loop characteristic has to be known to predict the stability of condition 1. The Nyquist loop for this transfer function shows that condition 1 is satisfied. A similar examination of the Nyquist plot, including the readily computed cable impedance shows that conditions 2 and 3 are satisfied. Thus the linear analysis indicates stable operation for the system of Fig. 1. This result was confirmed by tests of conditions 1, 2, and 3, individually and by the behavior of the system as a whole, both in the laboratory with a simulated power network for the cable, and in the final installation.

One of the most obscure aspects of the power-system behavior is that of equilibrium conditions after one or a series of large earth-potential disturbances. While the system is stable, in the sense that the transient due to a perturbation will disappear in a finite time once the disturbance has been withdrawn, the range of possible equilibrium positions (disregarding the overvoltage protective feature) is extremely wide, from perhaps 1,300 to 2,500 v at the cable terminals. The upper

Vol.
Pt. 1
1956

---

* The analysis, briefly summarized, was made by C. A. Desoer.

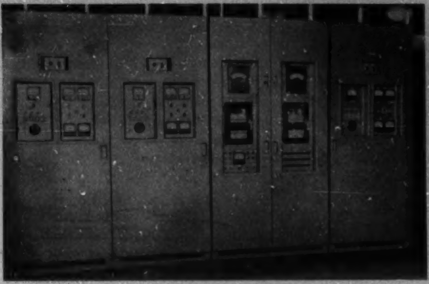


Fig. 10. D-c regulator and common control bays

limit of 2,500 v is set by the maximum output available from one power plant. This also sets the lower limit of the associate power plant at the far end of the cable. This situation is illustrated diagrammatically in Fig. 9.

The behaviors of the servomechanisms at the two ends of a given cable are nearly enough alike that the repeated introduction of simulated earth potential in the laboratory was found not to disturb substantially the equilibrium point. This was true for earth potentials of either polarity up to 1,000 v and with these potentials introduced at any point along the artificial cable. A rate of change of earth potential of 20 v per second was adopted in these tests, with the thought that such values would be realistic.

With regard to the long-term stability of the equilibrium condition as described, it is, of course, important that the controls which establish the cable current at the two ends of the same cable be adjusted for very nearly the same value. Unless this is done, the cable voltage at one end will gradually increase or decrease, and the voltage at the other end will move equally in the opposite direction. This would eventually bring in alarms and necessitate manual readjustment. In this connection, the voltmeters which indicate the drop through the series regulating tubes provide a very convenient magnification of any drift in cable current. The multiplying factor is the effective d-c impedance of the regulated system, i.e., more than 100,000 ohms. Thus 25 v, which is an appreciable fraction of the nominal 300 v across the series regulating tubes, is equivalent to less than 0.25 ma which is of the order of 0.1% of normal cable current. As a matter of fact, the behavior of this voltage provides the final criterion for precise adjustment of cable current to assure long-term stability.

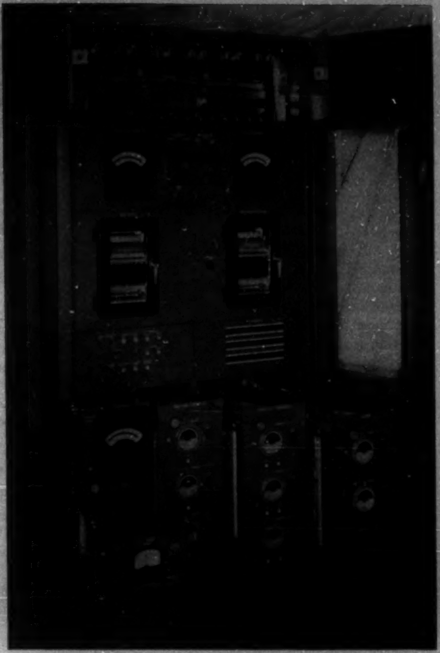


Fig. 11. Common control bay

Equipment Design

DESCRIPTION

Fig. 10 shows the complete d-c equipment for supplying one polarity of power to one cable. Similar equipment provides the opposite polarity to the other cable. The two equipments are located facing each other across a common aisle, with their common control bays directly opposite. Regulator 1, on the right, is normally operated in parallel with regulator 2. Regulator 3, on the left, is the spare regulator, and is normally off. The common bay, between regulators 1 and 2, includes the cable termination and power-separation filter. This equipment and all live parts of the circuit, back to the common point to which the switches in the individual regulator bays are connected, is enclosed in a high-voltage compartment.

The paralleling control switches are mounted in high-voltage compartments in their respective regulating bays and must remain completely enclosed, as their common cable connections are alive during cable operation. These switches have an interrupting capacity of 1.0 amp at 3,000 v, thus providing a large safety factor over the 0.245-amp maximum load current.

Fig. 11 illustrates some of the special design features built into the equipment to facilitate maintenance. The high-voltage compartment, shown open at the top, is locked whenever the cable is in operation, and this protection feature will be described. Pull-out drawers at the bottom contain metering shunts, a test unit for adjusting voltage and current-protection limits, a voltage-protection unit, a current-protection unit, and an alarm unit. While only one of these compartments is to be pulled out at a time, they are arranged so as not to endanger personnel or to affect service during adjustment when they are open. Doors are provided on all bays to prevent accidental disturbance of adjustments and to protect against damage to controls.

CORONA

The high-voltage a-c elements of the complete regulator bays were tested for corona with 4,000 v of alternating current applied. Furthermore, if corona was observed upon increasing the applied rms voltage to 5,000 v, it was required to extinguish when the voltage was reduced to 3,500 v. The maximum acceptable leakage was 20 μa (microamperes) at 4,000 v across the circuit (200 megohms). A d-c corona requirement of 4,000 v was applied to the d-c elements of the regulator bays and 5,000 v for the common bay, with a maximum permissible leakage of 5 μa. The higher corona requirements on the common bay were intended to eliminate the necessity for turning down the entire system for repair. A high standard of workmanship is required to provide such performance. There can be no sharp projections and no loose strands of wire. Solder must be applied in such a way as to obtain a rounded smooth joint, and high-voltage wiring must be dressed away from exposed grounded metal, bus bars, etc., so that the outer braid, other than polyethylene, does not come in contact with metal.

CROSSTALK AND OUTSIDE INTERFERENCE

To meet the severe crosstalk requirements between receiving and transmitting circuits and to guard against feeding office-noise potentials into the carrier transmission system, arrangements were made so that office grounds are carefully separated from the outer conductor of the cable and from all circuit elements within the power-separation filter. Pickup of external radio-frequency fields by the power-separation filters was greatly reduced by completely enclosing the cable terminal and the power-separation filter elements nearest to the terminal in a copper shield. The shielding itself and the cans of power-separation filter capacitors and oil-filled coils are connected to the return tape of the cable which is insulated from office ground until it reaches sea water, thus reducing the coupling to the other cable, as compared with tying both tapes together at the office or bay frame ground.

PROTECTION OF PERSONNEL

A key-locking system is provided to safeguard against any hazard to personnel from high voltages. In the common bay, the high-voltage compartment can be entered only by operating a switch which short-circuits the cable to ground and releases a key for the compartment doors. In each regulating bay, the key system assures that the bay is disconnected from the cable and hence from the paralleling power supplies. Where access is required to the interior of any compartment, the key system insures that the a-c power to the bay also is switched off.

The test compartment contains pin jacks, provided for maintenance operations, which are always performed with the regulator bay connected to a low-resistance load. Access to this compartment can be obtained with a-c power connected to the bay. However, for such access, the key system enforces the operation of the output disconnect switch, which also transfers the bay to a low-resistance load. Moreover, a mechanical interlock with the autotransformer assures that the test voltages are reduced to safe values.

Besides its function in protecting personnel, the key system also insures that no more than one regulating bay is disconnected at one time so that continuity of service is protected at all times by two parallel regulators.

Factory and Ship Cable Power

Besides the cable power supplies at the ocean terminals, similar d-c cable current-regulating equipment was designed for use at the cable factory and abroad the cable ship *Monarch*. Well-protected and closely regulated reliable power was considered essential during the cable loading and laying operations. It was necessary to have power on the cable continuously, except when splices were made, in order to detect a fault immediately, to measure transmission characteristics for equalization purposes, and finally to alleviate the strain on the glassware and tungsten filaments of the repeater tubes during the difficult laying period.[3]

References

1. SYSTEM DESIGN FOR THE NORTH ATLANTIC LINK, H. A. Lewis, R. S. Tucker, G. H. Lovell, J. M. Fraser. Page 850 of this issue.

2. REPEATER DESIGN FOR THE NORTH ATLANTIC LINK, T. F. Gleichmann, A. H. Lince, M. C. Wooley, F. J. Braga. Page 865 of this issue.

3. ROUTE SELECTION AND CABLE LAYING FOR THE TRANSATLANTIC CABLE SYSTEM, J. S. Jack, W. H. Leech, H. A. Lewis. Page 944 of this issue.

Electron Tubes for the Transatlantic Cable System

J. O. McNALLY G. H. METSON E. A. VEAZIE M. F. HOLMES

Synopsis: Electron tubes for use in repeatered underwater telephone cable systems must be capable of operating for many years with a reasonable probability of proper functioning. In the new transatlantic telephone cable system the section of the cable between Nova Scotia and Newfoundland contains repeaters developed by the British Post Office Research Station at Dollis Hill. These repeaters are built around the type-*6P12* tube developed at that research station. The repeaters contained in the section of the cable system between Newfoundland and Scotland are of Bell System design and depend on the *175HQ* tube developed at Bell Telephone Laboratories. In this paper the philosophy of repeater and tube design is discussed, and the fundamental reasons for arriving at quite different tube designs are pointed out. Some of the tube development problems and the features introduced to eliminate potential difficulties are described. Electrical characteristics for the two types and life test data are presented. Fabrication and selection problems are outlined and reliability prospects are discussed.

ELECTRON TUBES suitable for use in long submarine telephone cables must meet performance requirements that are quite different from those imposed by other communication systems. In the home entertainment field, for example, an average tube life of a few thousand hours is generally satisfactory. In the field of conventional land-based telephone equipment, where the replacement of a tube may require that a maintenance man travel several miles, an average life of a few years is considered reasonable. In deep-water telephone cables, such as the new transatlantic system, the lifting of a cable to replace a defective repeater may cost several hundred thousand dollars and disrupt service for an extended period of time. These factors suggest as an objective for submerged repeaters that the tubes should not be responsible for a system failure for many years, possibly 20, after the laying of the cable. Such

---

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957.

J. O. McNALLY and E. A. VEAZIE are with the Bell Telephone Laboratories, Murray Hill, N. J.; and G. H. METSON and M. F. HOLMES are with the British Post Office, London, England.

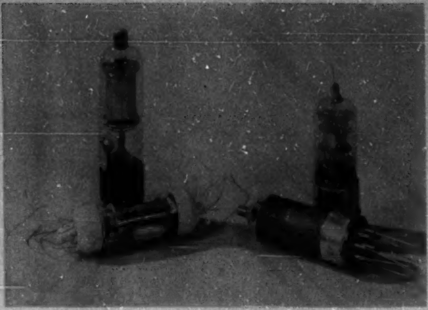


Fig. 1. Final designs of tubes for Nova Scotia–Newfoundland section of cable (right) and for Newfoundland-Scotland section (left). Early models of each type stand behind final models

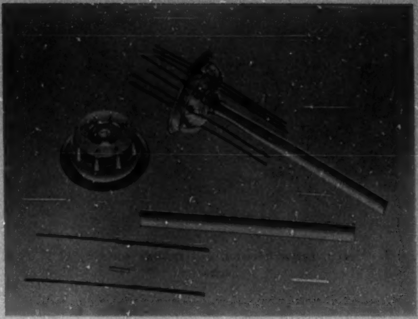


Fig. 2. Parts used in stem and a finished stem of the 175HQ tube. Separate beading of leads may be noted

very long life requirements make necessary special design features, care in the selection and processing of materials that are used in the tubes, unusual procedures in fabrication, detailed testing and long aging of the tubes, and the application of unique methods in the final selection of individual tubes for use in the submerged repeaters.

As indicated in the foreword and discussed at length in companion papers, the British Post Office developed the section of the cable system between Clarenville, Newfoundland, and Sidney Mines, Nova Scotia. This part of the transatlantic system used the *6P12* tube which was developed at the General Post Office Dollis Hill Research Station. The submerged portion of this system contains 84 tubes in 14 repeaters. Bell Telephone Laboratories developed the part of the system between Clarenville, Newfoundland, and Oban, Scotland. This section requires 102 repeaters, including 306 tubes, of a type known as the *175HQ*. Although a common objective in the development of the two sections has been to obtain very long life, the tube designs are quite different.

The Bell System decided on the use of a repeater housing that could be treated as an integral part of the cable, to facilitate laying in deep water. The housing is little larger than the cable and is sufficiently flexible to be passed over and around the necessary sheaves and drums. In such a housing the space for repeater components is necessarily restricted. This space restriction, combined with the general philosophy that the number of components should be held to an absolute minimum and that each component should be designed to have the simplest possible structural features, has resulted in the choice of a 3-stage 3-tube repeater. In this design, each tube carries the entire responsibility for the continuity of the service.

The Post Office Research Laboratories, prior to the development of the transatlantic cable system, had concentrated their efforts on shorter systems for shallow water. The placing of the repeaters on the bottom did not present the serious problems of deep-sea laying, so more liberal dimensions could be allowed for the repeater circuit. A 3-stage amplifier was developed which consisted of two strings of three tubes each, parallel-connected, with common feedback. The circuit was so designed that almost any kind of tube failure in one side of the amplifier caused very little degradation of circuit performance. This philosophy of having the continuity of service depend on two, essentially independent, strings of tubes has been carried over to the repeater design for the Clarenville–Sidney Mines section of the transatlantic cable.

In the Post Office system containing 84 tubes in the submerged repeaters, five tube failures randomly occurring in the system will result in slightly over 50% (per cent) probability of a system failure; one tube failure in the 306 tubes in the Newfoundland-Scotland section of the system will result in certain system failure. It is not surprising therefore to find the tube designed for the Newfoundland-Scotland section of the cable to have extremely liberal spacing between tube elements, to minimize the hazards of electrical short circuits. This results in a lower transconductance than is found in the tubes designed for the Nova Scotia–Newfoundland link. Other factors in the design will be recognized as reflecting the different operating hazards involved.

Early models of the British Post Office and Bell Laboratories tubes, together with the final tubes used in the cable system, are shown in Fig. 1.

Tubes for the Newfoundland–Scotland Cable

EARLY DEVELOPMENT CONSIDERATIONS

In Bell Telephone Laboratories, work on tubes for use in a proposed transatlantic cable was started in 1933. This was preceded by a study of what type of tube would best fit the needs of the various proposed amplifier systems, and by consideration of what might be expected to give the best life performance.

At the time this project was started, reasonably good tube life had been established for the filamentary types used in Bell System repeaters. Some groups of tubes had average lives of 50,000 or 60,000 hours (6 or 7 years). Equipotential cathode tubes were not then used extensively in the telephone plant, and there was no long-life experience with them. However, there appeared no basic reason why inherently shorter thermionic life should be expected, using the equipotential cathode, and there were several advantages in its use. One was the greater freedom in circuit design afforded by the separation of the cathode from the heater. Also there was the possibility of operating the heaters in series and using the voltage drop across the heaters for the other circuit voltages. It was felt, in addition, that the over-all mechanical reliability would be greater if the cathode were stiff and rigidly supported.

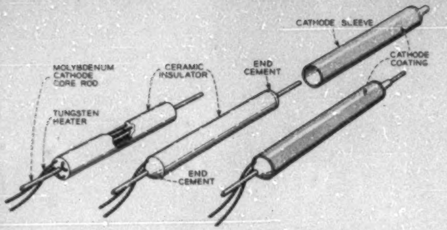


Fig. 3. Heater, heater insulator, and cathode assembly of 175HQ tube

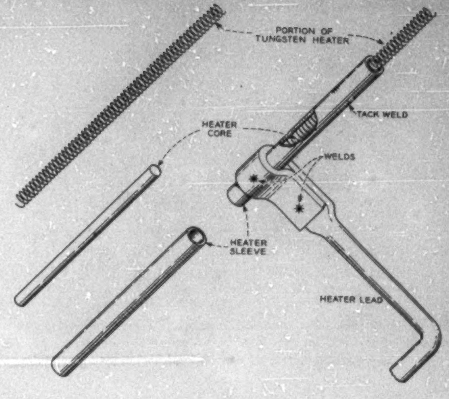


Fig. 4. Heater tabbing arrangement of 175HQ tube

The first equipotential tubes made were triodes. They were designed for use in push-pull amplifiers wherein continuity of service might be retained in case of a tube failure. This circuit was abandoned in favor of a 3-tube feedback amplifier that was the forerunner of the present repeater. The pentode was favored over the triode for this amplifier, for obvious reasons, and in 1936 the triode development was discontinued.

Early in the development of the tube, three basic assumptions were made. These were: 1. that operation at the lowest practical cathode temperature would result in the longest thermionic life, 2. that operating plate and screen voltages should be kept low, and 3. that the cathode current density should be kept as low as was practicable.

The first assumption, concerning the cathode temperature, was based on the observation of life tests on other types of tubes. While the data at the time of the decision were not conclusive, there was definite indication that too high a cathode temperature shortened thermionic life. Little was known about life performance in the temperature range below the values conventionally used.

The second assumption, concerning low screen and plate voltages, had not been supported by any experimental work available at the time of decision. Sixty volts were originally considered for the output stage; this value was later lowered when other operating conditions were changed. Subsequent results showed that in this range the voltage effects on thermionic life were relatively negligible.

The third assumption, that low-cathode current density favored longer thermionic life, affected the tube design by suggesting the use of a large, coated, cathode area. This implied the use of relatively high cathode power. It was decided, early in the planning of the repeater, that the voltage drop across the three heaters operated in series would be used to supply part or all of the operating plate and screen potentials. For a 60-volt plate and screen supply, the heater voltage could be as high as 20 volts. A quarter of an ampere was considered a reasonable cable current and consistent with voltage limitations at the cable terminals. Thus 5.0 watts were available for each cathode. With this power, a coated area of 2.7 square cm (centimeters) was provided. The value of the cathode current, the cathode area, and the interelectrode spacings define the transconductance. Very liberal interelectrode spacings were provided, consistent with reasonable tube performance. The original design called for a spacing of 0.040 inch between control grid and cathode. This value was later reduced to 0.024 inch, and a satisfactory design was produced which gave 1,000 micromhos or one ma (milliampere) per volt at a cathode current drain of approximately 2.0 ma. The resulting current density of approximately 0.7 ma per square centimeter is in sharp contrast with values such as 50 ma per square cm, used currently in tubes designed for the more conventional communication uses. Subsequent data, discussed later, indicate that for current densities of a few milliamperes per square centimeter, the exact value is not critical in its effect on thermionic life.

SUBSEQUENT PRODUCTION PROGRAMS

The development of the tube was pursued on an active basis through the years leading up to World War II. During the war, development activity essentially stopped. It was possible to keep only the life tests in operation. After the war the development of the tube was completed and a small production line was set up in Bell Telephone Laboratories, under the direct supervision of the tube development engineers, to make and select tubes for a cable between Key West, Fla., and Havana, Cuba.

This cable turned out to be a "field trial" for the transatlantic cable which was to come later. A total of six submerged repeaters containing 18 tubes was laid and the cable put in operation in June 1950. The cable has been in operation since this date without tube failure, and periodic observations of repeater performance indicate no statistically significant change in tube performance over the 6 years of operation.

Sufficient tubes were made, at the same time as the Key West-Havana run, for a future transatlantic cable. These tubes were never used, principally because they had been assembled with tin-plated leads. Tin plating, subsequent to the laying of the Key West-Havana cable, was found capable of growing "whiskers."[1]

In 1953, another production setup in Bell Telephone Laboratories was established for the fabrication of tubes for the Newfoundland-Scotland section of the transatlantic cable. On the completion of this job fabrication was continued in order to provide tubes for an Alaskan cable between Port Angeles, Wash., and Ketchikan, Alaska. After a pause of several months another run was made to provide tubes for a cable to be laid between California and the Hawaiian Islands.

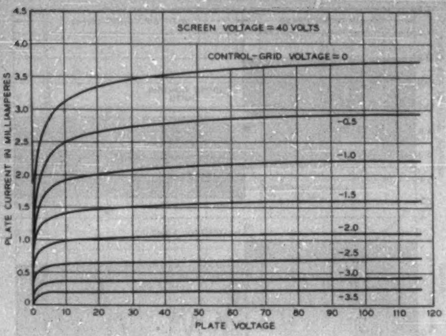


Fig. 5. Typical plate voltage-plate current characteristics for type 175HQ tube

Mechanical Features

The tube, shown on the left in Fig. 1, is supported in the repeater housing by two soft-rubber bushings into which the projections of the two ceramic end caps fit. All leads are flexible and made of stranded beryllium copper which has been gold-plated before braiding. Both for convenience in wiring in the circuit and for holding down the control-grid to anode capacitance, the grid lead has been brought through the opposite end of the tube from the other leads.

A number of somewhat unusual constructional features appear in the tube. The stem on which the internal structure is supported consists of a molded glass dish into which seven 2-piece beaded dumet leads are sealed. The parts used in a stem, and also a finished stem, are shown in Fig. 2. It is usual to embed the weld or "knot" between the dumet and nickel portions of the lead in the glass seal to provide more structural stiffness. This has not been done in this stem because it was believed that a fracture of a lead at the weld could be detected more easily if it were not supported by the seal. It might be questioned why the modern alloys and flat stem structure have not been used but it is to be remembered that one gas leak along a stem lead would disable the system. Experience gained with the older materials provides greater assurance of satisfactory seals.

The structure of the heater and cathode assembly is unique, as may be seen in Fig. 3. A heater insulator of aluminum oxide is extruded with seven holes arranged as shown. This insulator is supported by an 0.025-inch molybdenum rod inserted in the center hole. The heater, consisting of about 36 inches of 0.003-inch tungsten, is wound into a helix having an outside diameter of 0.013 inch. After dip coating by well-known techniques the heater is threaded through the six outer holes in the insulator. A suspension of aluminum oxide is then injected into the holes in the insulator so that on final firing the heater becomes completely embedded. The cathode sleeve, which is necked down at one end, as shown in Fig. 3, is slipped over the heater assembly and welded to the central molybdenum rod which becomes the cathode lead. By this means is obtained a uniform temperature from end to end of the cathode. Under normal operating conditions the heater temperature is approximately 1,100 C (degrees centigrade), which is considerably under the temperature found in other tubes.

Connection of the heater to the leads from the stem presented a serious design problem. Crystallization of tungsten during and after welding and mechanical strains developed by thermal expansion frequently are the causes of heater breakage. This problem was successfully overcome by the means illustrated in Fig. 4. Short sections of nickel tubing are slipped over the cleaned ends of the heater coil, and matching pieces of nickel wire are inserted as cores. These parts are held together by tack welds at the mid-points of the tubing. The heater stem leads are bent, flattened, and formed to receive the ends of the heater, which are then fastened by welds as indicated in the drawing.

A serious attempt has been made in the design of the tube to hold the number of fastenings that depend entirely on one weld to an absolute minimum. The grids are of conventional form in which the lateral wires are swaged into notches cut in the side rods or support wires. The side rods as well as the lateral wire are molybdenum. This produces grids which are considerably stronger than those for which more conventional materials are used.

The upper mica is designed to contact the bulb, and the bulb is sized accurately to receive and hold the mica firmly. The tube in its mounting will withstand a single 500*g* (gravitational acceleration) 1-millisecond shock without apparent changes in mechanical structure or electrical characteristics. It is estimated from preliminary laying tests that accidental or unusual handling would rarely result in shocks exceeding 100*g*.

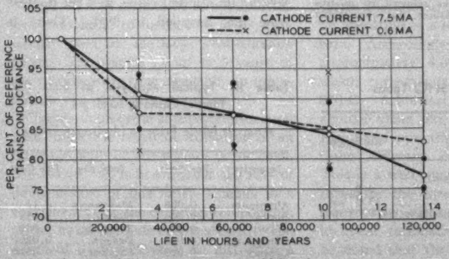


Fig. 6. Results of life tests on 18 175HQ tubes operating at two different current densities

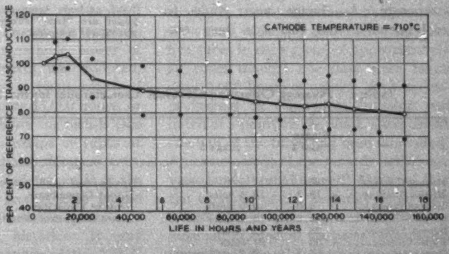
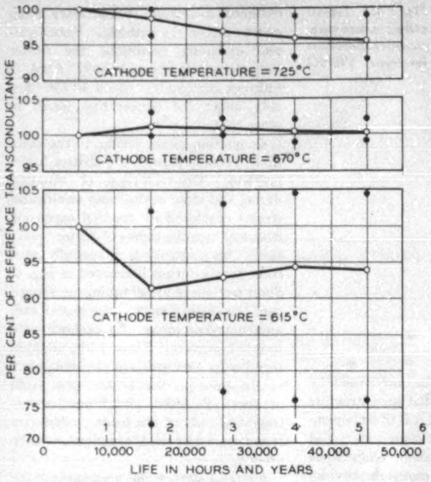


Fig. 7. Results of life tests on 16 175HQ tubes operating at cathode temperature of 710 C



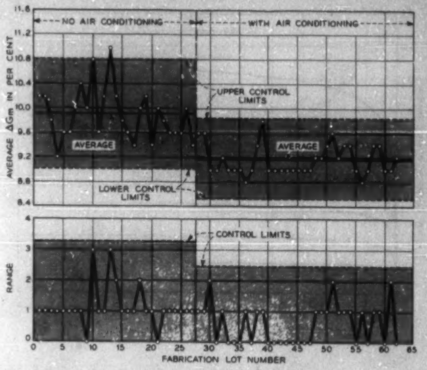


Fig. 8 (left). Results of operating 36 175HQ tubes divided equally among three different cathode temperature conditions. For each curve, cathode core material used was half from melt 60 and half from melt 63. Conditions in cable operation are essentially those represented by center curve

Fig. 9 (above). Control chart showing effect of air cleaning on cathode activity level. Per-cent change in transconductance for normal heater current and a value 20% lower used as measure of cathode performance

Electrical Characteristics and Life

The average operating electrical characteristics for the *175HQ* tube are given in Table I, and a family of plate-voltage versus plate-current curves for a typical tube is given in Fig. 5, for a region approximating the operating conditions.

The development of a long-life tube offers good opportunities to observe effects that are likely to be missed where shorter lives are satisfactory. For example, some of the earliest tubes made, after 20,000 hours on the life racks, began to show a metallic deposit on the bulbs. Immediate concern for the lowering of insulation resistance across mica spacers prompted an investigation. The source of the trouble was traced to the use of plates made from a grade of nickel which allowed magnesium as a contaminant to evaporate. A change was made to molybdenum, and this has been employed successfully since that experience.

The effect on the thermionic life of operating at different cathode current densities was of interest. Life tests were started in which the cathode current drain in one group of tubes was approximately 7.5 ma (2.8 ma per cm^2) and in another group the average cathode current was 0.6 ma (0.2 ma per cm^2). The test results presented in Fig. 6 after 120,000 hours, or the equivalent of approximately 14 years, show that at the cathode temperature of approximately 710 C* selected for the test, there is practically no current density effect in this 12-to-1 current range. The circles indicate average values, while the dots and crosses at each test point show the positions of the extreme tubes of the group.

Similar life tests, set up to show the effect of operating at plate and screen voltages of 60 volts as compared to 40 volts, indicate no essential differences in performance after 8 years of operation.

The cathode temperature is one of the most critical operating variables affecting thermionic life. As already mentioned, the early development objective was a cathode power 5.0 watts which corresponded to 710 C for the cathode design used at that time. The results of operating at this condition are illustrated in Fig. 7. No tubes have been lost from the test where the direct cause has been failure of emission. Several tubes were lost because of mechanical failure resulting from design defects which were subsequently corrected. It will be observed that at the end of 17 years the average transconductance is 80% of its original value, and the poorest tube has dropped to 69%. There is

---

* All cathode temperatures referred to in this paper are "true" temperatures, not uncorrected pyrometer temperatures.

Table I. Average Operating Electrical Characteristics for the 175HQ Tubes

	Stages 1 and 2	Stage 3
Heater Current	220	217 milliamperes
Heater Voltage	18.2	18.4 volts
Heater Power	4.0	4.0 watts
Control-Grid Bias	−1.3	−1.4 volts
Screen Voltage	38	40 volts
Plate Voltage	32	51 volts
Screen Current	0.3	0.3 milliamperes
Plate Current	1.3	1.4 milliamperes
Transconductance	980	1,010 micromhos

	All Stages
Capacitances (cold, with shield)	
Input Capacitance	9.2 μμf
Output Capacitance	15.6 μμf
Plate to Control-Grid Capacitance	0.03 μμf

Table II. Typical Analysis of Inco 220 Nickel Cathode Melt 84

Analysis Made Prior to Hydrogen Firing

Impurity	Per Cent	Impurity	Per Cent
Aluminum	0.008	Manganese	0.11
Boron	<0.004	Silicon	0.033
Cobalt	0.46	Titanium	0.032
Chromium	<0.005	Oxygen	0.0001
Copper	0.028	Sulphur	0.0016
Iron	0.093	Carbon	0.058
Magnesium	0.046		

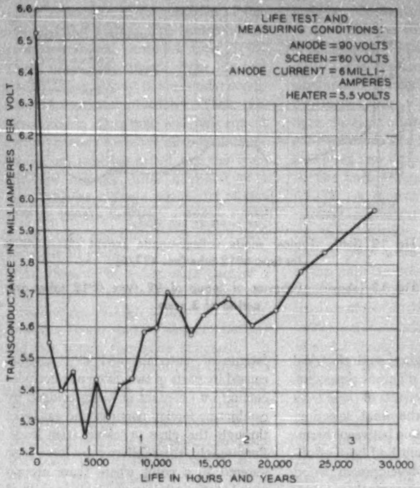


Fig. 10 (left). Behavior of group of 50 tubes deliberately left with "gas generator" (tube type 6P12)

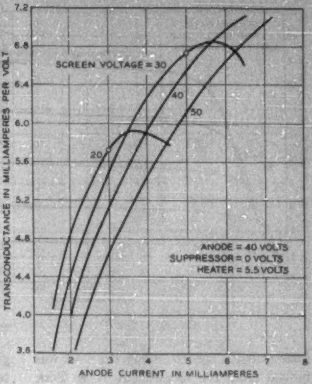


Fig. 11 (right). Typical transconductance-anode current characteristics for type 6P12 tube (no. 457/6)

reason to believe that test-set difficulties may very well account for a large part of the variation shown in the first 3 years.

The cathode coatings used in all experimental and final tubes for the Newfoundland-Scotland link of the transatlantic cable are the conventional double-carbonate coatings. The cathode base material is an International Nickel *220* nickel. The particular melt used for the transatlantic cable is known as melt *84*. A typical analysis for melt-*84* nickel cathodes is given in Table II.

The relatively high carbon content (0.058%) of melt *84*-cathode nickel is capable of producing excessive reduction of barium in the cathode coating.[2,3] A treatment in wet hydrogen prior to coating, at 925 C for 15 minutes, reduces the carbon in the cathode sleeve to about 0.013%.

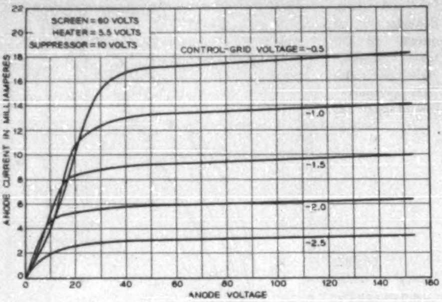
Melt *84* was as close in composition as was obtainable to melts *60* and *63*, previously used for the Key West–Havana tubes. The results of up to 5 years of life testing were thus available on materials of very similar composition.

One common cause of tube deterioration with life is the result of formation of an interface layer on the surface of the cathode sleeve. It is known that the rate of development of this layer depends in a complex way on the chemical composition of the nickel cathode core material. The effect of such a layer is to introduce a resistance in series with the cathode. This results in negative feedback and reduces the effective transconductance. Since the effect of a given feedback resistance in this location is proportional to transconductance, the relatively low value for the *175HQ* tube tends to minimize this feedback effect. In addition, the low cathode temperature tends to reduce the rate of formation of interface resistance, and the relatively large cathode area tends further to minimize the effects. The final decision to use melt *84* was based on accelerated aging tests which showed it to be superior to melts *60* and *63* from an interface standpoint. (The interface problem will be discussed further in a later section.)

As the development of the tube proceeded, both the processing of the parts and the cleanliness of the mount assembly were improved, and the cathode emission level increased. Life tests indicated that better thermionic life might be obtained by operating at a lower cathode temperature. Accordingly, a cathode power of approximately 4.0 watts was adopted, which corresponds to a temperature of 670 C. A life test, now 45,000 hours, or about 5 years old, shows the results (see Fig. 8) of operating groups of tubes at three different cathode temperatures. This is a well-controlled test in that the tubes for the three groups were picked from tubes having common parts and identical fabrication histories. It may be noted that the average of the 725 C lot has lost approximately 5% of the initial transconductance, whereas the 4.0-watt group after about 5 years has lost essentially none of its transconductance. The 3.0-watt (615 C) group shows serious instabilities in its performance. In some of the tubes the cathode temperature has not been sufficiently high to provide the required emission levels.

The design of the repeaters in the Newfoundland-Scotland section of the cable is such that reasonably satisfactory cable performance would be experienced if the transconductance in each tube dropped to 65% of its original value. The life test performance data presented in Figs. 7 and 8, and other tests not shown, indicate that operation of the *175HQ* tubes in the transatlantic cable at approximately 4.0 watts will assure satisfactory thermionic performance for well over 20 years.

Mention was made that cleanliness in the assembly of the mounts was a factor which affected thermionic activity. Interesting evidence supporting this view was obtained during the fabrication of tubes for the Key West–Havana cable. The quality control type of chart reproduced in Fig. 9 shows the average change in transconductance between two set values of heater current for the first five tubes in each group of approximately 24 tubes made. The data were taken after 5,000 hours of aging. A sharp improvement in thermionic emission was noted at a point on the chart where about one half of the tubes had been fabricated. An examination of the records (which are very carefully maintained) disclosed that the windows of the assembly room were sealed and air cleaning and conditioning were put in effect at the point indicated on the chart. No other changes in processing or materials occurred at this time. A second definite improvement in therm-



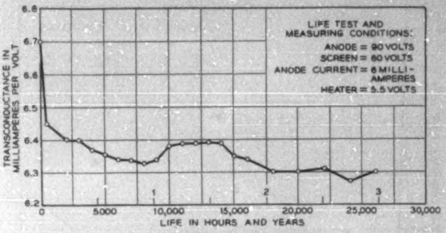


Fig. 12 (left). Typical anode voltage–anode current characteristics for type 6P12 tube (no. 457/6)

Fig. 13 (above). Behavior of group of 92 type 6P12 tubes over period of 3 years

ionic emission occurred when the work was moved from the location in New York City to the new and better controlled environment at Murray Hill, N. J.

Fabrication and Selection

All assembly operators on the *175HQ* tube program wore nylon* smocks to keep down the amount of dust and lint that might otherwise leave their clothing and get into the tubes. Rayon acetate gloves were worn in handling parts as a protection against perspiration. Rubber finger cots did not prove satisfactory because they covered too little area, and once contaminated they did not absorb the contaminant.

The tubes for the Newfoundland-Scotland section of the cable were made at Bell Telephone Laboratories under the extremely close supervision of many of the original development engineers. All materials going into the tubes were carefully checked, and wherever possible they were tried out in tubes. Experience under accelerated aging conditions was obtained before these materials were used. For example, although during the developmental period all glass bulbs were used as received without any failures resulting, less than one quarter of the bulbs bought for the actual cable job passed the inspection requirements. Each batch of heaters was sampled and results were obtained on intermittent and accelerated tests before approval for use.

The fabrication of the tube was carried out with extreme care by operators specially selected for the job. If normal commercial test limits had been applied to the tubes after exhaust, the yield from acceptable mounts would have been about 98% to 99%. Yet only about one out of every seven tubes pumped was finally approved for cable use. All tubes were given 5,000 hours' aging, and electrical tests were made at six different times during this period. The results of these tests weighed heavily in the final selection. For example, a correlation between thermionic life and gas current had been established during the tube development period and only tubes having control-grid currents caused by gas of less than 5×10^{-11} ampere were acceptable. This corresponds to a gas pressure of approximately 2×10^{-7} mm (millimeters) of mercury. Very thorough mechanical inspections were made after the 5,000 hours' aging to ensure that there were no observable mechanical deviations that could cause trouble. The history of each group of 28 tubes, from which prospective candidates for the cable were selected, was reviewed to see if any group abnormalities were found. In case a suspected trait was seen, all tubes in the group were ruled out for cable use.

As an aid in the selection of tubes for the cable, all pertinent data were put on International Business Machine cards. It was then possible to manipulate and present the data in many very helpful ways that would have otherwise been wholly impractical from the standpoints of time and man-power considerations. An over-all total of about half a million bits of information was involved.

Reliability Prospects

Questions are frequently raised concerning the probability of tube failures in the system. There are two categories into which failures naturally fall: catastrophic failures, and failures caused by cumulative effects, such as the decay of thermionic activity, development of primary emission from the control-grid, or the build-up of conductance across mica insulators or glass stems.

The catastrophic failures include open connections caused by weld failures or fatigue of materials, and short circuits caused by parts of two different electrodes coming into contact or being bridged by conducting foreign particles and gas leaks through the glass or along stem leads. Fortunately, these failure rates have been lowered to a point where there are no sound statistical data available in spite of the substantial amount of life testing that has been done. In approximately 4,800 tubes made to date, there have been four failures that were not anticipated by the inspection made. The four failures were of different types, and occurred either at or before 5,000 hours of life. All four were of types more likely to occur during the early hours of aging and handling.

Among the cumulative types of failure, life testing has indicated no apparent problem with either the growth of insulation conductance or primary emission from the grids. As indicated earlier in the paper, thermionic life results are such that there is reason for optimism that no failures will occur in 20 years.

Tubes for the Nova Scotia–Newfoundland Cable

Trend of Tube Development in British Submarine Repeater Systems

Early Use of Commercial Receiving Tubes

The development of submerged telephone repeaters in Britain has taken a somewhat different course from that followed in the United States. Off North America, deep seas are encountered as soon as the continental shelf has been passed. Consequently, emphasis has been placed from the beginning on the design of repeaters for ocean depths. In Britain, separated from many countries by only shallow seas, it was natural to start with a repeater specially designed for shallow water. Such a repeater was laid in 1944 in an Anglo-Irish cable.

---

* Trade name for DuPont polymide fiber.

The tubes used in the amplifier of this repeater were normal, high-transconductance, commercial pentodes, type *SP61*. These tubes were known, from life test results, to last at least 2 years under conditions of continuous loading. Their performance in the first and subsequent early repeaters exceeded all expectations. So far, one tube has failed, and this from envelope fracture, after a period of 4 years' service. There remain 23 of this type on the sea bed with a service life of 5 or 6 years, and three tubes which have survived 10 years.

All these *SP61* tubes were part of a single batch made in 1942. Their performance set a high standard; it was, however, found that subsequent batches did not attain the same standard. In 1946 therefore the British Post Office was faced with the fact that future development of the shallow-water system of submerged repeaters was dependent on the production of a tube type which could take the place of the 1942 batch of *SP61* tubes. This situation led to the formation of a team at Dollis Hill whose terms of reference were, specifically, to produce the replacement tube and, generally, to study the problems presented by the use of tubes in submerged repeaters. Apart from changes in the specific requirements, these terms of reference have remained unchanged from that day to this.

Replacement by the G.P.O. 6P10 Type

Coincident with the rapid exhausting of stocks of satisfactory *SP61* tubes for submerged telephone systems, there arose the need for a tube type for a submerged telegraph repeater. This latter requirement was complicated by the fact that the telegraph cable was subject to severe overall voltage restrictions which precluded the 630-ma heater current required for the 4-watt cathode of the *SP61*. To avoid production of one tube type for telephone systems and another for telegraph, it was decided that the replacement for the *SP61* should have a 2-watt cathode with a 300-ma heater.

During 1944 and 1945, a very successful, miniature, high-slope pentode, the *CV138*, was produced for the armed services. Its electrical characteristics were superior to those of the *SP61* and, in addition, it used a 2-watt cathode. It was therefore decided to base the replacement tubes, electrically, on the *CV138*, while at the same time retaining freedom to amend the mechanical features in any way which might seem to favor the specific requirements of submerged repeater usage: in particular, maintenance of the level of transconductance unchanged for long periods. Consequently, three major mechanical changes were made at the outset of the project. The miniature bulb of the *CV138* was replaced by one of normal size (approximately 1 inch in diameter and $2^1/_2$ inches long). This was done to reduce the glass temperature and so reduce gas evolution. At the same time, a normal press and drop seal were substituted for the button base and ring seal of the *CV138*, as it was felt that, with the techniques available, the former would be the more reliable. With the use of a normal press there immediately followed a top cap control-grid connection, and thus was produced a double-ended tube in place of the single-ended *CV138*.

These three modifications and a number of major changes to improve welding and assembly techniques led to the General Post Office type known as the *6P10* (a pentode *P* with a 6.3-volt heater *6* of design mark *10*). The *6P10* replaced the *SP61* in the 18 shallow-water repeaters laid in various cables after 1951. There are therefore 54 type *6P10* tubes in service on the sea bed with periods of continuous loading ranging from 2 to 4 years. There has been one failure due to a fractured cathode tape, and one other repeater was withdrawn from service to investigate a high-frequency oscillation associated with a tube. The oscillation cleared, however, before the cause could be identified.

The first 18 of the *6P10*-type tubes used in repeaters had conventional nickel cathode cores. Appreciation of the problem of interface resistance led to the use of platinum as a core material for the following 36 tubes. The steps leading to this radical change of technique will be described later.

Development of the 6P12 for Long-Haul Systems

Although submerged-repeater development started naturally in Britain with shallow-water systems, it was inevitable that attention should ultimately turn towards transoceanic cables. The tube requirements for such long-haul systems differ from those for short-haul shallow-water schemes, in that operation at a lower anode voltage is essential. A new tube to replace the *6P10* was therefore unavoidable.

By the time emphasis started to shift in Britain from shallow-water to deep-sea systems, considerable experience had been gained at Dollis Hill on the production techniques required to fit a *6P10*-type tube having a platinum cathode core for submerged-repeater usage. When it became apparent that a new tube had to be designed for the first long-haul system, it was resolved to retain as much as possible of the *6P10* structure in order to take full advantage of familiar techniques. The *6P10* was therefore redesigned for 60-volt operation simply by a major adjustment to the screen-grid position and minor alterations elsewhere. The new tube became known as the *6P12*, and was used in seven repeaters installed in the Aberdeen-Bergen cable. This scheme was regarded as a proving trial for the Newfoundland–Nova Scotia section of the transatlantic project.

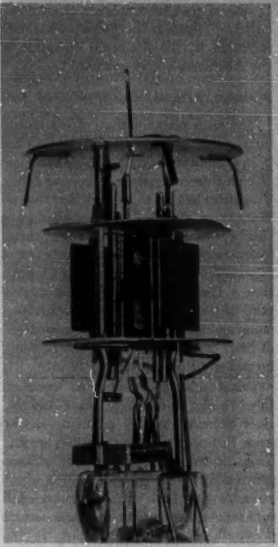


Fig. 14. Interior of 6P12-type tube

It has always been appreciated that the use of high-transconductance tubes using closely spaced electrodes will involve a higher liability to mechanical failure by internal short circuits. Practical experience in shallow-water schemes, where repeater recovery is a comparatively cheap and simple operation, seems to show, however, that such risks are outweighed by the economic advantage of a tube capable of wider frequency coverage. In point of fact, a failure by internal short circuit has not yet occurred on any shallow water system.

This background of experience explains the British choice of a high-transconductance tube for deep-sea systems, but the greater liability to mechanical failure by use of parallel amplifiers is acknowledged. Confidence in this policy has been increased by the successful operation of the Aberdeen-Bergen system.

Problems of Development of the 6P12 Tube

The main preoccupation of the thermionics group at Dollis Hill since 1946 has been a study of the electrical life processes of high-transconductance receiving tubes. This effort has led to a conviction that all changes of electrical performance have their origin in chemical or electrochemical actions occurring in the tube on a micromicro or millimicro scale of magnitude. The form of change most important to the repeater engineer is decay of transconductance; this will be considered in brief detail as typical of the developmental effort put into the *6P12* tube.

Transconductance decay in common tubes results from two separate and distinct chemical actions occurring in the oxide cathode itself. Both actions are side issues, in no way essential to the basic functioning of the cathode, and it seems probable that both can be eliminated if sufficient understanding of their nature is available. The first action is the growth of a resistive interface layer between the oxide matrix and its supporting nickel core (discussed briefly in an earlier section). This effect is assumed to be due to silicon contamination of the nickel core metal, as follows:

$$BaO + Si = Ba_2SiO_4 + 2\ Ba$$

The resistance of the layer of barium orthosilicate rises as it loses its barium activator and approaches the intrinsic state. The effect of the interface resistance is to bring negative feedback to bear on the tube with resulting loss of transconductance. The second deleterious action is loss of electron emission from the oxide cathode by direct destruction of its essential excess barium metal by oxidizing action of residual gases. Such gases result from an imperfect processing technology.

These two problems have been approached in the *6P12* tube in a somewhat novel manner. The conventional nickel core is replaced by platinum of such high purity (99.999%) that the possibility of appreciable interface growth from impurities can be disregarded. The only factor to be considered is the appearance of high-resistive products of a possible interaction between platinum and the alkaline earth oxides. Batch tests over a period of 30,000 hours have failed to show any sign whatever of such an action occurring, and workers at Dollis Hill now regard the pure platinum-cored tube as free from the interface resistance phenomenon.

The problem of avoiding gas deactivation of the cathode is a more difficult one and so far has been reduced rather than eliminated. It is now appreciated that the dangerous condition arises from "gas generators" left in the tube, and not from a true form of residual gas pressure after seal-off from the pump. These gas generators are solid components of the tube which give off a continual stream of gas over a prolonged period of time. The gas evolution rates are usually so small that they cannot be detected by reverse-grid current measurement, but they tend to integrate gas by absorption on the cathode and to destroy its activity. The gas generators are usually of finite magnitude and, depending mainly on diffusion phenomena, evolve gas at a rate which falls in roughly exponential fashion with time. The probability of transconductance failure is therefore highest in early life and tends to lessen as the generators move to exhaustion.

One particularly useful feature of the platinum-cored cathode is its freedom from core oxidation during gas attack, which leaves the tube free to recover from transconductance failure when the gas attack has passed. Fig. 10 shows the behavior of a group of 50 tubes which have been deliberately left in possession of a component capable of generating carbon monoxide over a prolonged period. The curve shows the characteristic recovery of a platinum-cored oxide cathode, with the gradual passing of what is thought to be a typical gas attack.

One problem that has attracted much attention at Dollis Hill is the actual manner in which a platinum-cored cathode recovers from a gas attack. The mechanism must involve the dissociation of a small fraction of the oxide cathode itself, with the retention of barium metal in the oxide lattice, and the evolution of oxygen. That such an essential mechanism does in fact exist has been proved by the slow accumulation of barium metal in the platinum core. This accumulation takes the form of a distinctive alloy of barium and platinum, and occurs only when the cathode is passing current. The barium regenerative process therefore seems to be electrolytic in nature and, depending only on current flow and a stock of oxide, would appear to be virtually inexhaustible.

These few remarks are perhaps sufficient to give some idea of the lines on which the British research effort has run during the past decade. More detailed descriptions have already been presented elsewhere.[4–7]

ELECTRICAL AND MECHANICAL CHARACTERISTICS

Electrical Characteristics

The main electrical characteristics of the *6P12* are shown in Figs. 11 and 12. The heater voltage used for both sets of curves is 5.5 volts, the same value as that used in the British amplifier.

Fig. 11 shows the change of transconductance with anode current, with screen voltage as parameter. An anode voltage of 40 volts and a suppressor voltage of zero correspond with the static operating conditions in the first two stages of both the Aberdeen-Bergen and the British T.A.T. (transatlantic telephone) amplifiers. The screen voltage and anode current of the first two stages of the British T.A.T. amplifier were chosen to be 40 volts and 3 ma respectively.

Fig. 12 shows the normal anode voltage–anode current characteristics for conditions corresponding to the output stage of the amplifier (static operating point, anode voltage = 90, screen voltage = 60, anode current = 6 ma). A final electrical characteristic worthy of comment is the level of reverse-grid current. For all specimens tested at the time of selection, after about 4,000 hours' life test the level is very low, about 100 micromicroamperes per milliampere of anode current.

Life Performance

The life performance of the *6P12* is still a matter for conjecture. The only concrete evidence available is the behavior of a group of 92 tubes placed on life test some 3 years ago. The change of the average transconductance of this group (with anode current constant at 5 ma) is shown in Fig. 13. It may be well appreciated that there is no definite trend over the past year which permits any firm prediction of life expectancy. Examination of other tube characteristics is equally unproductive from the point of view of prediction of failure.

In the early stages of the test there were eight mechanical failures. The cause in all instances was identified and corrected in subsequent production before the start of the T.A.T. project.

Mechanical Characteristics

The chief mechanical characteristics of the *6P12* have been referred to as very similar to those of the *6P10*. A photograph of the interior of the tube is shown in Fig. 14.

TUBE SELECTION TECHNIQUES

Not all the tubes that were found after production to be potentially suitable for

Vol.
Pt. 1
195

the British T.A.T. amplifier, remained equally suitable after the life test period of about 4,000 hours. A brief account of how the best were selected follows.

The fact that every tube had to pass conventional static specification limits needs little emphasis here. This test was, however, supplemented with three additional types of specification. First, every tube was tested in a functional circuit, simulating that stage of the amplifier for which the tube was ultimately intended. Here measurements were made of shot noise (appropriate to first-stage usage), and harmonic generation (appropriate to the output stage), in addition to the usual measurements of transconductance, anode impedance, and working point.

Second, all tubes were subject to intensive visual scrutiny in which some 80 specific constructional details were checked for possible faulty assembly.

Third, the life characteristics of transconductance, total emission and working point were examined over the test period of about 4,000 hours for unsatisfactory trends. Although this type of specification is more difficult to define precisely, its application is probably more rigorous and exacting than any of the previous specifications.

Only if a tube passed the conventional test and the three supplementary tests was it considered adequate for inclusion in a repeater.

Conclusions

The laying of the present repeatered transatlantic cable represents by far the most ambitious use to date of long-life, unattended electron tubes. In this project alone there are 390 tubes operating on the ocean bottom. If to this number are added the ocean bottom tubes from earlier, shorter systems, those used in the Alaskan cable completed a few months ago, and those to be used in the California-Hawaii cable to be laid in 1957, the total number on the ocean bottom will be about 1,000. The capital investment dependent on the satisfactory performance of these tubes is probably about 100 million dollars—strong evidence of faith in the ability to produce reliable and trustworthy tubes.

It is of interest to note that the two groups working on the tubes on opposite sides of the Atlantic had no intimate knowledge of each other's work until after the tube designs had been well established. As a result of subsequent discussions, it has been surprising and gratifying to find how similarly the two groups look at the problems of reliability of tubes for submarine cables.

The authors would be completely remiss if they did not mention the contributions of others in the work just described. These projects would have been impossible had it not been for the enthusiastic, cooperative and careful efforts of many people working in varied fields. Over the years chemists, physicists, electrical and mechanical engineers, laboratory aides, shop supervisors, and operators all have made essential contribution to the projects. It would be impractical and unfair to attempt to single out for mention the work of specific individuals whose contributions are outstanding. These are too many.

References

1. Filamentary Growths on Metal Surfaces—"Whiskers," K. G. Compton, A. Mendizza, S. M. Arnold. *Corrosion*, Houston, Tex., vol. 7, Oct. 1951, pp. 327–34.

2. The Influence of Impurities in the Core-Metal on the Thermionic Emission from Oxide-Coated Nickel, M. Benjamin. *Philosophical Magazine and Journal of Science*, London, England, vol. 20, July, 1935, p. 1.

3. Initial Emission and Life of a Planar-Type Diode as Related to the Effective Reducing Agent Content of the Cathode Nickel (abstract only), H. E. Kern, R. T. Lynch. *Physical Review*, New York, N. Y., vol. 82, May 15, 1951, p. 574.

4. The Life of Oxide Cathodes in Modern Receiving Valves, G. H. Metson, S. Wagener, M. F. Holmes, M. R. Child. *Proceedings*, Institution of Electrical Engineers, vol. 99, pt. III, Mar. 1952, p. 69.

5. Deterioration of Valve Performance Due to Growth of Interface Resistance, G. H. Metson, M. F. Holmes. *Post Office Electrical Engineers' Journal*, London, England, vol. 46, Jan. 1954, p. 193.

6. The Growth and Properties of Cathode Interface Layers in Receiving Valves, M. R. Child. *Ibid.*, vol. 44, Jan. 1952, p. 176.

7. A Study of the Long Term Emission Behaviour of an Oxide Coated Valve, G. H. Metson. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 102, pt. B, Sept. 1955, p. 657.

Cable Design and Manufacture for the Transatlantic Submarine Cable System

A. W. LEBERT H. B. FISCHER M. C. BISKEBORN

Synopsis: The transatlantic cable project required that two repeatered cables be laid in the deep-water crossing between Newfoundland and Scotland, and one across the shallower waters of Cabot Strait. The same structure was adopted for the cables laid in the two locations.

This paper discusses the considerations leading to design of the cable and describes the method of manufacture, the means and equipment for control of cable quality, the process and final inspection procedures, the electrical characteristics of the cable, and factors relating to mechanical and electrical reliability of the final product.

Description of Cable

GENERAL features of the cable structure adopted for the transatlantic cable project[1] are illustrated in Fig. 1. The cable consists of two basic parts: 1. the coaxial, or the electrical transmission path, and 2. the armor or outer protection and strength members.

The coaxial is made up of three parts: 1. the central conductor, 2. the insulation, and 3. the outer or return conductor. The central conductor is composed of a copper center wire surrounded by three helically applied copper tapes. The insulation is a polyethylene compound which is extruded tightly over the central conductor. The insulated central conductor is called the cable core. The outer or return conductor is composed of six copper tapes applied helically over the insulation.

The protection and strength components shown in Fig. 1 for the type *D* deep-water cable are provided by a teredo tape of thin copper applied over the outer coaxial conductor, a fabric tape binding, a layer of jute rove for armor bedding, the textile-covered armor wires and, finally, two layers of jute yarn flooded with an asphaltum-tar compound. This type of cable is characterized by the extra tensile strength of its armor wires and the precautions taken to minimize their corrosion.

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957.

A. W. Lebert, H. B. Fischer, and M. C. Biskeborn are with Bell Telephone Laboratories, Murray Hill, N. J.

The authors wish to acknowledge the many contributions made by the members of the staff of the Simplex Wire and Cable Company, Submarine Cables Limited, the British Post Office, and the Bell Laboratories groups involved during the course of the design and manufacture of the transatlantic cable.

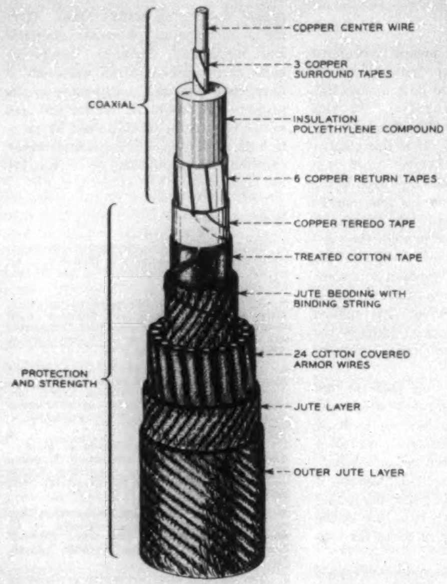


Fig. 1. Structural features of deep-water type of cable

At the shallow-water shore ends, the armor types are characterized by the use of mild steel wires which are increased in diameter in steps as the landing is approached. These types are designated *A* and *B* and will be described in greater detail later.

The transmission loss of this cable structure at the top operating frequency of 164 kc is 1.6 db (decibels) per nautical mile and is 0.6 db per nautical mile at 20 kc, which is the lower end of the frequency band. The high-frequency impedance of the cable is about 54 ohms.

Basis of Design

A coaxial structure was first used for telephone and telegraph service in a submarine installation in 1921, between Key West and Havana, Cuba. Three coaxial cables with continuous magnetic loading and no submerged repeaters were laid. One telephone circuit and two telegraph circuits were provided in each cable for each direction of transmission.

In 1950, a pair of submarine coaxial cables,[2] which included flexible submerged repeaters, was laid between Key West and Havana, Cuba. Each cable furnished 24-voice circuits. One cable served as the "go" and the other as the "return" for the telephone conversations. The transatlantic telephone cable design is similar to this cable except that the nominal diameter of the insulation is 0.620 inch instead of 0.460. An outstanding difference between the transatlantic and Key West–Havana systems is cable length, about 2,000 nautical miles as compared with 125. This difference influenced significantly the permissible electrical and mechanical tolerances applying to the cable structure.

The installation of some 1,200 miles of cable with island-based repeaters for a communication and data transmission system for the U. S. Air Force, between Florida and Puerto Rico,[3] followed the 1950 submarine cable system. The design of this cable is identical with that of the transatlantic cable, except for differences in the permissible dimensional tolerances on the components of the electrical transmission path. Data obtained on the electrical performance of the Air Force cable provided the transmission characteristic to which the repeaters for the transatlantic project were designed.

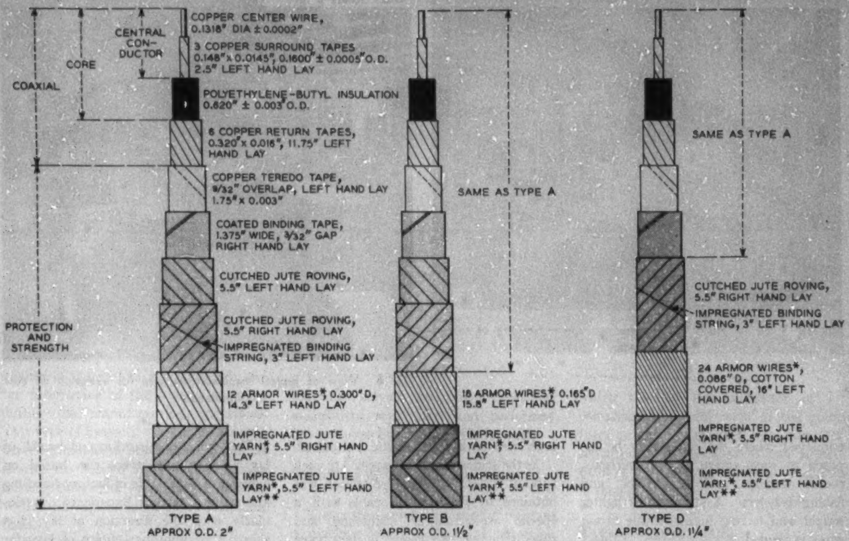
The design of these cable installations were the result of many years of cable development effort, which was guided by the successes and failures of the earlier submarine telegraph cables. The 1950 Key West–Havana and the Air Force cables differed from the earlier structures in one important respect, namely, the lay of the major components. A series of fundamental design studies during the 1930's and 1940's and extensive field tests in the Bahamas in 1948 demonstrated that having the same direction of lay of the major components of a cable was very important in minimizing kinking and knuckling. In addition, other laboratory tests pointed the direction for the adoption of new materials and techniques in the manufacture of these cables. These and subsequent improvements in materials and manufacturing techniques were included in the transatlantic cable design.

Since the electrical characteristics of a cable have a direct bearing on the over-all system design and performance, considerable emphasis was placed on this phase of the design of the transatlantic cable. The size of central conductor and core used in the Air Force cable resulted in low unit attenuation and low d-c resistance. These advantages resulted in the adoption of this size of cable. However, the outside diameter of the core and the diameter of the central conductor of the coaxial do not fulfill the requirements generally described as optimum for minimum attenuation. Mathematical analysis shows that there is a preferred diameter ratio which results in minimum transmission loss. For the 0.620-inch core diameter employed in the Air Force cable, the central conductor diameter chosen was smaller than the ideal central conductor required to satisfy the preferred diameter ratio. The diameter chosen retained the central conductor size which the Key West–Havana and Air Force cables had proved to be satisfactory from a manufacturing standpoint. The choice was also compatible with the d-c resistance requirement for transmission of power over the cable to each of the repeaters.

While production of the cable was proceeding, cable manufactured to the transatlantic specification was tested near Gibraltar in March 1955. These tests provided a final evaluation of the mechanical and electrical characteristics of this cable before the actual laying of the transatlantic link.

Details of Structural Design of Cable

The structural features of the coaxial and of types *A*, *B*, and *D* armor are summarized in Fig. 2. A composite


COPPER CENTER WIRE, 0.1318" DIA ± 0.0002"
CENTRAL CON-DUCTOR
3 COPPER SURROUND TAPES 0.148" x 0.0145", 0.1600" ± 0.0005" O.D. 2.5" LEFT HAND LAY
CORE
COAXIAL
POLYETHYLENE-BUTYL INSULATION 0.620" ± 0.003" O.D.
6 COPPER RETURN TAPES, 0.320" x 0.016", 11.75" LEFT HAND LAY
COPPER TEREDO TAPE, 3/32" OVERLAP, LEFT HAND LAY 1.75" x 0.003"
COATED BINDING TAPE, 1.375" WIDE, 3/32" GAP RIGHT HAND LAY
CUTCHED JUTE ROVING, 5.5" LEFT HAND LAY
PROTECTION AND STRENGTH
CUTCHED JUTE ROVING, 5.5" RIGHT HAND LAY
IMPREGNATED BINDING STRING, 3" LEFT HAND LAY
12 ARMOR WIRES*, 0.300" D, 14.3" LEFT HAND LAY
IMPREGNATED JUTE YARN*, 5.5" RIGHT HAND LAY
IMPREGNATED JUTE YARN*, 5.5" LEFT HAND LAY**
TYPE A
APPROX O.D. 2"
SAME AS TYPE A
18 ARMOR WIRES*, 0.165" D 15.8" LEFT HAND LAY
IMPREGNATED JUTE YARN*, 5.5" RIGHT HAND LAY
IMPREGNATED JUTE YARN*, 5.5" LEFT HAND LAY**
TYPE B
APPROX O.D. 1½"
SAME AS TYPE A
CUTCHED JUTE ROVING, 5.5" RIGHT HAND LAY
IMPREGNATED BINDING STRING, 3" LEFT HAND LAY
24 ARMOR WIRES*, 0.086" D, COTTON COVERED, 16" LEFT HAND LAY
IMPREGNATED JUTE YARN*, 5.5" RIGHT HAND LAY
IMPREGNATED JUTE YARN*, 5.5" LEFT HAND LAY**
TYPE D
APPROX O.D. 1¼"

*HOT FLOODING COMPOUND APPLIED DURING THESE OPERATIONS
**COATED WITH CALCIUM CARBONATE

Fig. 2. Design details of types of transatlantic telephone cable for different depths of water

central conductor was chosen to provide a conductive bridge across a possible break in any one of its elements, due to a hidden defect, such as an inclusion of foreign material in the copper. The dimensions of the components of the central conductor were precisely controlled, and a light draw through a precision die was used to compact and size the assembly.

Use of high molecular weight polyethylene (grade 0.3) for core insulation is a major departure from the materials used in early submarine telephone and telegraph cables. The development of synthetic polymers such as polyethylene had led to the replacement of gutta percha as cable insulation, since polyethylene possesses better dielectric properties and mechanical characteristics and is lighter in weight.

Ordinary polyethylene of low molecular weight is subject to environmental cracking, especially in the presence of soaps, detergents, and certain oils. Material of high molecular weight is much less subject to cracking, and by adding 5% (per cent) butyl rubber, further crack resistance is obtained.

Six copper tapes applied helically over the core comprised the return conductor and thus completed the coaxial structure. The dimensions of these tapes were precisely controlled. The helical structure was chosen to impart flexibility to the coaxial.

Insulation of some of the early submarine telegraph cables suffered from attack by marine borers such as the teredo, pholads, and limnoria. To protect against such attack, a thin metallic tape was placed over the insulation in the early submarine cables. The necessity for such protection for the transatlantic cables, especially in deep water, may be questioned, but the moderate cost of this protection was considered cheap insurance against trouble. The copper teredo tape was applied directly over the return conductor, as a helical serving with overlapped edges to completely seal the coaxial from attack by all but the smallest marine organisms.

A cotton tape treated with rubber and asphaltum-tar compound was applied over the teredo tape to impart mechanical stability to the coaxial during manufacture. A small gap between adjacent turns of the helix was specified to permit ready access of water to the return tape structure and to the surface of the core. The use of a gap was based on laboratory tests which showed that transmission loss was dependent to a modest extent on thorough wetting of the exterior of the coaxial. Since transmission loss measurements are made on repeater sections of cable shortly after manufacture to determine whether any length adjustments are required, it was essential that the wetting action be as rapid as possible.

The design of the protection and strength components of the cable was modified according to the depth of the water in which the cable was to be laid. To prevent damage to the coaxial by any cutting action of the armor wires during manufacture and laying, a resilient cushion of jute roving was placed between the armor wires and coaxial. For type *D* cable, a single layer of jute was used; for types *A* and *B*, the bedding was made up of two layers of jute. To protect this jute from microbiological attack, a cutching treatment was employed. The traditional process consists of treating the jute with a vegetable compound called catechu or cutch.

Armor wires were applied over the bedding jute. The use of heavy or intermediate weight near shore has been established by experience with ocean cable. This type of armor is generally

Fig. 3. View of typical manufacturing area for stranding of central conductor

Fig. 4. View of typical manufacturing area for extrusion of core insulation

employed where the cable may be exposed to wave action, bottom currents, rocks, icebergs, ship's anchors, and fishing trawlers. A structure of lighter weight and having higher tensile armor wires is needed in deep water. Table I shows the essential differences between the armor types employed in the transatlantic cable and the approximate range of depths in application.

In addition to these armor types, a shore length of 0.6 nautical mile was provided with an insulated lead sheath under type-*A* armor to facilitate preferred grounding arrangements and to provide signal-to-noise improvement.

Table I

	Armor Wire			
Type	Number of Wires	Diameter, Inches	Material	Application, Fathoms
A	12	0.300	mild steel	up to 350
B	18	0.165	mild steel	350 to 700
D	24	0.086	high-strength steel	greater than 700

Where the tensile strength of the armor wires is most important, as in the type-*D* design, each of the wires was protected against corrosion by a zinc galvanize plus a knitted cotton serving or helically applied tape, the whole assembly being thoroughly saturated with an asphaltum-tar compound. The effectiveness of such protection is clearly apparent when early submarine cables, which used this protection, are recovered and examined. For the heavier armor types, the protection was similar to that of type *D*, except that the textile serving was replaced by a dip treatment to coat each wire with an asphaltic compound.

As the armor wires were applied to each type of cable, additional protection was obtained by flooding the cable with a special asphaltum-tar compound and then applying two layers of jute yarn over the wires. The jute yarn was impregnated with an asphaltum-tar compound before application to the cable and then flooded with another asphaltum-tar compound after application. Formulation of cable flooding materials required the use of compounds having a relatively high coefficient of friction to avoid slippage of the cable on the ship's drum during laying.

To assure satisfactory handling characteristics during the laying operation, all of the metallic elements of the cable were applied with a left-hand direction of lay and the lengths of lay (except for the teredo tape) were chosen so that approximately the same helical length of material was used per unit length of cable. Since the teredo tape was relatively soft and ductile compared to that of the other metallic components, it was not necessary to equate its helical length to that of the other components. Width and lay of the teredo tape were selected to give a smooth, tight covering.

The choice of direction and length of lay of the jute layers was based on experience with cable in factory-handling and laying trials. Experience, particularly with the direction of lay, has shown that improper choice of lays for the two outer layers of jute may result in a cable that is difficult to coil satisfactorily in factory and ship storage tanks. The combination of lays selected for the cable components provided good performance in all the handling operations, including the final laying across the Atlantic.

Manufacture of the Cable

Before considering the manufacture of the cable, it should be understood that the repeater gain characteristic was designed to compensate for the loss characteristic of the cable. Therefore, once this loss characteristic was established, it was essential that all cable manufactured conform with this characteristic.

To obtain the required high degree of conformance, close control had to be kept over all stages of manufacture and over the raw materials. Controls to guide the manufacture of the cable were set up with two broad objectives:

1. To produce a structure capable of meeting stringent transmission requirements.
2. To assure that the manufactured cable could be laid successfully and would not be materially affected by the ocean bottom environment for the expected life of the cable system.

Attainment of a final product capable of meeting the stringent transmission

Fig. 5 (left). Special water tank or pan for tests on immersed cable core

Fig. 6 (right). View of shop for application of cable armor

requirements is described in a later section. Process and raw material controls in manufacture were provided by an inspection team which checked the quality of the various raw materials and the functioning of the several processes during the manufacturing operations. This type of inspection coverage is somewhat unique with submarine cable. It assures the desired final quality by permitting each error or accident to be investigated and corrected on an individual basis.

Cable for the transatlantic crossing was manufactured in the United States by the Simplex Wire and Cable Company and in England by Submarine Cables, Ltd. Differences in machinery and equipment in the plants of the two manufacturers necessitated minor differences in the sequence of the operations and in the processes. The sequence of operations in assembly of the cable was as follows:

Step No.	Operation
1	stranding of central conductor
2	extrusion of insulation
3	runover examination, repair where necessary
4	panning and testing of core
5	jointing of core
6	application of return tapes, teredo tape, fabric tape, jute bedding, and binding string
7	application of armor wire and outer jute layers
8	storage in tanks, testing
9	splicing in repeaters, testing

The only important difference in the sequence of the manufacturing operations at the two plants was the use of separate operations for steps 6 and 7 at Submarine Cables and the combination of these operations in one machine at Simplex.

Other minor differences in process methods related to raw materials. For example, the American supplier purchased polyethylene already compounded with butyl rubber and antioxidant in granule form, ready for use. The British supplier purchased polyethylene, butyl rubber, and antioxidant separately, and performed the compounding in the cable factory.

Stranding of Central Conductor

The central conductor was stranded on a machine which included a revolving carriage with suitable arbors for the three surround tapes. It was equipped with brakes designed to assure equal pay-off tension among the tapes and with detectors to automatically stop the strander in case of a tape break. Each tape was guided through contoured forming rolls to shape the tape to the center wire.

The joints between successive reels of wire and tape used in fabrication of the central conductor were butt brazed. The brazes were staggered to avoid more than one braze in a given cross section of the conductor. The quality of the brazes in these components was controlled by a qualification technique described in the section "Jointing of Core Lengths."

The strand was drawn through several forming dies to size the finished diameter of the central conductor accurately. No lubrication was used because the removal of the resultant residues, which could contaminate the polyethylene insulation, was difficult. The taper in the central conductor diameter due to the die wear was controlled by appropriate replacement of tungsten carbide dies, where used, or by the use of a diamond die where the rate of die wear is less than 1 or 2 micro-inches per mile.

The stranding area in both plants was enclosed and pressurized to guard against dirt and dust settling on the central conductor. A high standard of cleanliness was maintained for parts of the machine which touched the conductor or its components. Undue wear of the guide faces or capstan sheaves was cause for replacement of the sheaves and adjustment of the machine. The stranding area is shown in Fig. 3.

Extrusion of Insulation

To avoid possible contamination of the polyethylene insulating compound in the extruder-hopper loading area, a pressurized enclosure prevented entry of airborne dust and dirt, and the containers of polyethylene compound were cleaned with a vacuum cleaner before being brought into the hopper area. A fine screen pack placed in the extruder reduced the possibility of contamination in the core.

In passing through the extruder, the central conductor was payed off of a large reel with controlled tension, into the pay-out capstan, through an induction heater, through a vacuum chamber, and thence into the cross-head of the extruder. The induction unit heated the central conductor and provided means for controlling the shrinkback of the core insulation and the adhesion of the conductor to the insulation. Shrinkback is a measure of the contained stresses in the insulation.

On the output side of the extruder, the core was cooled in a long sectionalized trough containing progressively cooler water from the input to the output end. The annealing of the polyethylene in the cooling trough also served to hold the shrinkback of the core to a low value. A view of the extrusion shop is shown in Fig. 4.

An important addition to the extrusion operation consisted of the use of an im-

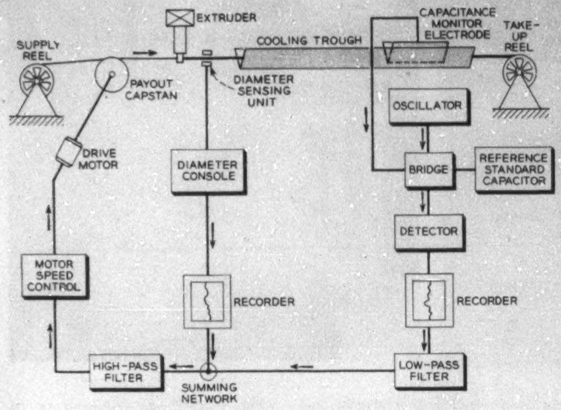


Fig. 7. Simplified block diagram of capacitance monitor servo-control system

proved servo system to control the extruder automatically for attaining constant capacitance per unit length of coaxial. The system used is described in a subsequent section.

Runover Examination

Following extrusion, the core was subjected to continuous visual and tactual examination in a rereeling operation called "runover." The purpose of the runover operation was the detection of inclusions of foreign material in the dielectric and the presence of abnormally large or small core diameters. Core not meeting specification requirements was cut out or repaired.

In addition to visual inspection of the cable, examination of short lengths of core was made at regular intervals with a shielded source of light arranged to illuminate the interior of the dielectric material. Provision of this internal illumination facilitated detection of particles of foreign material well beneath the surface of the core. Strip chart records of the unit-length capacitance of the core obtained during extrusion were used as a guide in searching out regions of uncertainty.

Panning and Testing

After runover, the core was coiled in tanks of water, as shown in Fig. 5. Precautions were taken to remove the air dissolved in the water and thus prevent the formation of bubbles on the surface of the core. The water was also temperature-controlled and circulated to maintain uniform temperature throughout the tank. Thermocouples placed at different levels in the tank determined when the temperature was uniform. Measurements of d-c conductor resistance, a-c capacitance, insulation resistance, and dielectric strength were then made. These measurements are discussed in detail in a later section.

Jointing of Core Lengths

The cable core was manufactured in lengths much shorter than a repeater section, which necessitated connecting the individual core lengths together. Jointing techniques consisted of brazing the central conductor with a V-notch type of junction and of molding in a short section of the polyethylene insulation. After silver-soldering the V-joint, a safety wire consisting of four fine-gauge tinned copper wires was bridged across the junction in an open helix and soft soldered at the ends. Extreme care was taken to remove any excess rosin and to eliminate any sharp points on the ends of the safety wires. The safety wire is intended to maintain continuity of the electrical path in case the braze should fail.

Visual examination of brazes in the actual cable was the only means available for their final inspection. To assure a high degree of quality on these brazes, a system was devised for checking the performance of the operator and the brazing machine initially and at frequent intervals through the use of sample brazes in each of the components, which were tested to destruction. To control the uniformity of brazes, the brazing of the copper wire and tapes was made as automatic as possible by the use of controlled pressure on the components, appropriate-sized wafers of silver solder, and an automatically timed heat cycle. The tests on the brazes used to qualify operators and machinery indicated that a high degree of braze performance was achieved.

Pressure and temperature were carefully controlled during the molding of the insulation over the conductor. Periodic checks similar to those described for brazes were made on operator and molding machine to maintain a satisfactory level of performance. In addition, each molded joint placed in the actual cable was X-rayed and subjected to a high-voltage test while immersed in water.

Application of Return Tapes and Armor Wires

After the core lengths were joined together, they were pulled through the return taping and armoring operations. The machine for applying return tapes was designed specifically for the purpose and was similar in characteristics to the corresponding portion of the strander for the central conductor. Controlled pay-off tension, automatic breakage detectors, precision guides, and contoured forming rolls to shape the tape, were incorporated in the construction.

The return tape, teredo tape, and fabric tape were applied from taping heads, and the bedding jute and binding string were applied from serving heads in a tandem operation. Another set of tandem operations included the application of armor wires, outer jute layers, and the appropriate asphaltum-tar flooding compounds. In the American supplier's plant, both sets of tandem operations were combined into one continuous production line. In the British plant, these operations were divided into two separate production lines. A view of the armoring machine area is shown in Fig. 6. Following the application of the flooding compounds, whiting is applied either at the take-up capstan on the armoring line or in the storage tanks as the cable is coiled.

To avoid core damage, the flow of hot flooding compound was stopped when the cable in the armoring line was stopped. One of the major sources of such stoppages was the reloading of the various applicating heads.

Storage and Testing

In a continous haul-off operation, the cable was conveyed from the armoring machine to the tank house for storage. The cable was coiled in spiral layers, called flakes. Each flake started at the outside rim of the tank and worked toward the central cone. Several 37-nautical-mile repeater sections were stored in each tank.

Water was circulated through the cable tanks to establish uniform temperature conditions throughout the mass of cable. When thermocouples located at appropriate points in the tank indicated that the cable temperature was uniform, measurements were made of attenuation, internal inpedance irregularities, and terminal impedances, d-c resistance, d-c capacitance, insulation resistance, and dielectric strength.

To facilitate these tests without interrupting production, successive repeater section lengths were placed in alternate tanks. By this procedure, a group of four or five sequential repeater section lengths, called an "ocean block," was stored in two tanks. The ends of each repeater section were brought out of the tank to a splicing location. After all tests were completed and the specification requirements met, the repeaters were spliced in. Testing of the ocean block for transmission performance completed the manufacturing operations.

Raw Materials

Stringent requirements were placed on all raw materials used in the manufacture of the transatlantic cable. Detailed specifications covered the basic requirements and the methods of controlling their quality on a sampling inspection procedure. The requirements for the materials were established to ensure that their use would not jeopardize the life of the cable. Since cable life is critically related to the integrity of the insulation, all materials had to be scrutinized for their tendency to cause environmental cracking. These tests were necessarily made on an accelerated basis. Since no correlation exists at present between accelerated tests and long-term (20-year) life tests, only conservative design selections can be justified.

Close tolerance such as ±0.0002 inch for the diameter of the solid center wire in the central conductor were specified for all copper components of the coaxial. In addition, these components had to be free from slag or other inclusions, and the wire drawing and rolling of the tape had to be controlled to assure smooth surfaces, edges of prescribed shape, and freedom from filamentary imperfections. Compounds used in drawing and rolling operations were selected to minimize the possibility of contaminating or causing cracking of the polyethylene. Residual quantities of compound on the wire or tapes were removed prior to annealing, which was controlled to prevent the formation of oxides and to assure clean and bright copper.

The dielectric constant range of the polyethylene-butyl compound was limited to 2.25 to 2.29. These limits were determined by the limited accuracy of the measuring equipment available at the time. Restrictions covered the allowable amount of contamination, since its presence in other than minute quantities might have reduced the dielectric strength or degraded the power factor of the compound.

In addition, the melt index (a factor related to molecular weight) of the final insulating compound composed of polyethylene resin, butyl rubber, and antioxidant was held to 0.15 to 0.50. The melt index of ordinary polyethylene, used for insulation generally, is 2.0 or higher. The low index assured the maximum resistance to environmental cracking.

The cutching and fixing processes used in the manufacture of bedding jute were adjusted to limit the alkalinity of the jute because of the adverse effect of alkaline materials on polyethylene compound. Oils used in the spinning of the jute were selected to obtain types which were not strong cracking agents for polyethylene, and the quantities used were reduced to the workable minimum. The presence of such impurities as bark and roots was restricted to provide the desired fiber strength. The impregnation of the jute was controlled to ensure adequate distribution of the coal tar throughout the fibers without having an excess that would make the jute difficult to handle during the armoring process.

The size, composition, and processing of the armor wires were also placed under close control. Purity, tensile strength, and twist requirements were designed to ensure that the wire could be applied to the cable, and welded, and that it could withstand the expected tensions during laying and pickup. Strength considerations made it mandatory that inclusions of slag or piping of the wire be eliminated. Piping is an unusual condition encountered during rolling or drawing, which results in a hollow shell of steel which may be filled with slag.

Fig. 8. Photograph of capacitance monitor electrode and servo-controller console in laboratory setup

Control of Transmission Characteristics

From a broad point of view, the attainment of a final product capable of meeting the stringent transmission requirements was achieved by the following basic steps.

1. Precision control of the dimensions of the copper conductors, including the diameter of the fabricated central conductor.

2. Automatic control of the insulating process to maintain a constant capacitance, thus compensating for deviations in central conductor diameter and dielectric constant of the insulation.

3. Factory process control, by means of a running average of the measured attenuation characteristic of current production, to guide the adjustment of suitable parameters when necessary.

As indicated in the sections concerning the method of manufacture and the control of raw materials, precautions were taken to obtain a central conductor that had predictable electrical performance, and a controlled taper in over-all diameter along its length. The need for such effort is explained by consideration of the factors that determine the attenuation of a coaxial structure.

The attenuation α of the cable is directly proportional to the a-c resistance R and inversely proportional to the characteristic impedance Z_0 as a satisfactory approximation, i.e.

$$\alpha = \frac{aR}{Z_0}\ db/nm$$

where a is coefficient depending on the units. It is thus clear that control of α may be attained by control of R and Z_0. Since the resistance is a function largely of the diameter of the central conductor, and since it is held to close tolerances, the constancy of impedance completes the requirement for attenuation control.

The characteristic impedance of a transmission line is determined by the following:

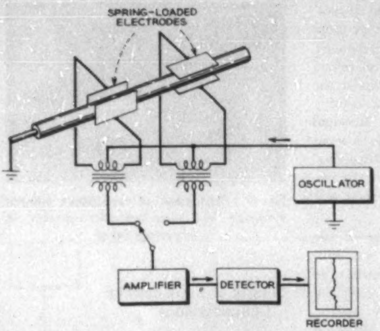


Fig. 9. Simplified block diagram of concentricity gauge for continuous measurement of centering of central conductor

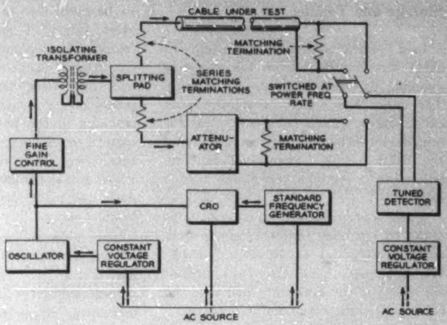


Fig. 10. Simplified block diagram of cable attenuation measuring set

$$Z_0 = b\sqrt{\epsilon}\log\frac{D}{d}\text{ ohms}$$

where ϵ is the dielectric constant of the insulating material, D is the inside diameter of the outer conductor, d is the diameter of the inner conductor, and b is a numerical coefficient. If the dielectric constant of the insulating material (polyethylene) does not vary, control of characteristic impedance reduces to control of capacitance. This follows from the fact that the capacitance C is related to the D/d ratio as follows

$$\frac{D}{d} = \text{antilog}\frac{k\epsilon}{C}$$

where k is a numerical coefficient.

Precision control of capacitance during the insulating process is achieved by a double-loop linear servo system, as shown in the simplified block diagram, Fig. 7. The two loops consist respectively of one capable of introducing relatively fast capacitance corrections of only modest accuracy, and of one capable of highly precise capacitance control on a relatively long time basis. The servo system controls the capstan payout speed of the central conductor feeding into the extruder applying the core insulation, as shown in the block diagram. Since the extruder delivers insulating material at a constant rate, an increase in central conductor speed results in a thinner-than-normal wall of insulation and thus causes an increase in capacitance.

The sensing element for the control loop consists of a capacitance monitor. This is a device capable of measuring the unit length coaxial capacitance of the cable core continuously as it moves through the water in the trough. Since the capacitance of a polyethylene-insulated core is temperature-sensitive, the monitoring electrode must be located at a point in the cooling trough where the temperature of the core is stable and known to a degree commensurate with the over-all accuracy objectives. The distance from the extruder to the electrode corresponds to about 10 minutes of cooling time; hence, a servo system based on this loop would be necessarily slow, because of the 10-minute delay in detecting a drift in capacitance.

Analysis shows that fast capacitance information of only moderate accuracy may be used in combination with the slow loop to speed up the response of the over-all system to a satisfactory degree, without sacrifice of precision of the slow loop. The sensing element used for the fast loop consisted of a light-ray diameter gauge, which measures the diameter (changes in diameter are the approximate inverse of the capacitance) of the hot core close to the extruder. The data for slow and for fast are combined to control the extruder, as shown in the block diagram.

The servo constants were chosen to minimize the deviations in unit length capacitance occurring in core lengths corresponding to less than one-fourth wave length of the top operating frequency. In other words, the objective for choice of servo loop constants was to assure equality in the capacitance of all one-fourth wave sections of core. Echo measurements indicated that a highly satisfactory degree of control was achieved. Over-all servo system performance was such that the standard deviation of the capacitance of the core lengths manufactured for the two crossings was ±0.1%. The capacitance monitor electrode and the servo console are illustrated in Fig. 8.

Adjustment of Concentricity

Means for setup and adjustment of the extrusion process to achieve relatively accurate centering of the conductor in its sheath of insulation was provided by a device called a concentricity gauge. This device operates on the principle that two small plane electrodes on opposite sides of the core will have different direct capacitances to the central conductor, when the conductor is not properly centered.

A simplified block diagram of the concentricity gauge is shown in Fig. 9. Data obtained with two sets of electrodes displaced 90 degrees were recorded on a strip chart recorder, with the output of the two sets of electrodes being displayed alternately. A satisfactory degree of centering was moderately easy to maintain.

Electrical Measurements

To assist in achieving the goal of matching the cable and the repeater characteristics with a minimum of deviation, electrical measurements were made throughout the process and close tolerances were placed on the electrical parameters in each stage of production. Measurements on the repeater section lengths of cable were used as a final check to determine the extent to which all of the controls had been successful. The primary standards used were calibrated by the Bureau of Standards in the United States or the National Physical Laboratories in England. These precision standards were used to calibrate the bridges frequently.

The d-c resistance of the central conductor was measured under constant

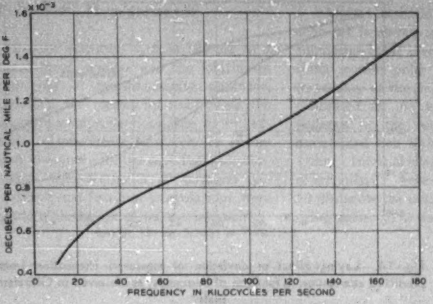


Fig. 11. Change in cable attenuation due to temperature as function of frequency

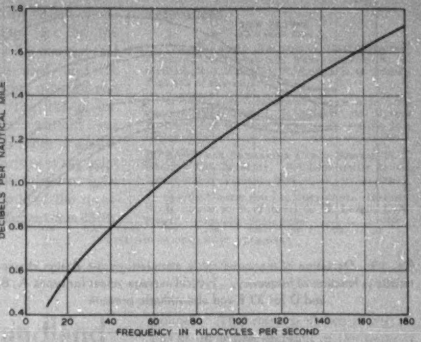


Fig. 12. Design characteristic of cable attenuation as function of frequency for 37 F and atmospheric pressure

temperature conditions with a precision type of Wheatstone bridge. The permissible range of resistance was 2.514 and 2.573 ohms per nautical mile at 75 F (degrees Fahrenheit). In practice, the spread of resistance values was well within these limits.

The 20-cycle core capacitance was also measured under constant-temperature conditions. For this measurement, the two ends of the central conductor were connected together and the measurements made between the central conductor and ground, which was provided by the water. A capacitance-conductance bridge was used for this purpose. The capacitance limits set initially were from 0.1726 to 0.1740 microfarad per mile at 75 F. Analysis of the core measurements shows that at each factory the range of capacitance was held more closely than indicated, which illustrates the benefits of servo control to the insulating process.

The d-c insulation resistance of the core was also measured by applying 500 volts for one minute. A minimum insulation resistance requirement of 100,000 megohm-miles at 75 F was established, but any lengths that had less than 500,000 megohm-miles were scrutinized for possible sources of trouble and were subject to rejection. As a general rule, insulation resistances considerably in excess of 500,000 megohm-miles were obtained.

The core was tested also at a voltage of 90,000 volts direct current for a period of one minute. This test was designed to catch any gross faults in the core caused by foreign particles which escaped detection by the other mechanical and electrical tests made on the core.

As discussed in the section on jointing, the core lengths were assembled and joined together to form a repeater section of cable. In general, an effort was made to produce the core for a repeater section of cable on a particular strander, extruder, and armoring line, and to join the lengths together in the order of manufacture. Practical difficulties such as the fact that the outputs of two stranders were required to supply one extruder made it impossible to achieve this objective in all cases.

Capacitance deviations from the desired nominal resulted from a variety of causes, such as inaccurate control of the temperature of the water in the core-cooling troughs and improper adjustment of the control apparatus. To minimize the reflection which would result from joining together two lengths of core of widely different capacitances, cores were not joined together if their measured a-c capacitances differed by more than 0.3%. When such capacitance differences did exist, the core length involved was removed from its normal sequence and placed in a position near the middle of the repeater section.

Because the taping and armoring processes were combined in one production line in the American factory, no other electrical measurements could be made on the components of the cable until it was completely armored. In the British factory, tests for information purposes only were made on the cable in the coaxial stage. These tests included measurements of attenuation, internal impedance irregularities, and terminal impedances. They served as a means of evaluating the changes in the electrical performance during armoring.

The insulation resistance requirement after armoring and storage under water for at least 24 hours was 100,000 megohm-miles. In addition, the cable had to withstand 50,000 volts for a period of one minute without failure.

Attenuation Measurements

As an aid in achieving the desired uniformity of product, new measuring equipment of improved accuracy was provided. A block schematic of this equipment is shown in Fig. 10. The requirements for this equipment were that it should be capable of measuring a 37- to 44-mile section of cable with an absolute accuracy of 0.04 db and a precision of 0.01 db in the frequency range from 1 to 250 kc.

The attenuation of the cable was measured at 10-kc intervals from 10 kc to 210 kc and measured values were corrected to 37 F, using the changes in attenuation owing to temperature, shown in Fig. 11. By comparing the corrected values with the design characteristic shown in Fig. 12, the deviations were determined. Both the attenuation characteristic and the changes in attenuation owing to temperature were derived from factory measurements of attenuation made on the Florida–Puerto Rico cable.

By comparing the running average and spread of these deviations with the design requirements, it was possible to assess the performance of the cable and, if required, to make any necessary adjustments in parameters for subsequent sections. In addition, these deviations were used to determine the length adjustment required for each repeater section to keep the sum of the deviations at each frequency in any one ocean block

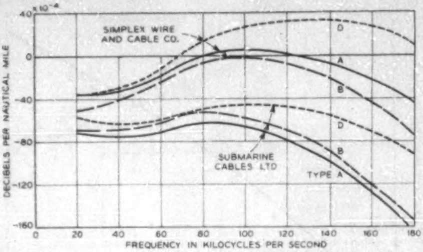


Fig. 13. Deviation of measured cable attenuation from design characteristic as function of frequency. Typical average values for types A, B, and D for 37 F and atmospheric pressure

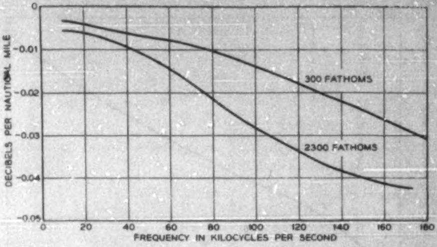


Fig. 14. Laying effect or deviation of measured attenuation from predicted attenuation as function of frequency, as observed in Gibraltar trials

to a minimum. Typical average attenuation deviation characteristics are shown in Fig. 13.

Temperature and Pressure Coefficients

Measurements of primary constants were made on 20-foot lengths of cable and core placed in a temperature-controlled and pressure-controlled tank. These measurements were used to compute the temperature coefficients of attenuation in order to check the values derived from measurements made on the Florida–Puerto Rico cable section. Additional attenuation measurements were made on several repeater section lengths of cable over a range of temperature from approximately 40 to 70 F, to establish further the magnitude of the changes in attenuation with temperature. The measurements indicated that the derived temperature coefficients were accurate to within ±10%.

Measurements also were made to determine the effect of pressure on the primary constants of the cable. These measurements indicated that capacitance was the only parameter affected by pressure. The capacitance increased linearly 0.1% for each 500 pounds per square inch of applied pressure. Since the attenuation α is inversely proportional to the impedance, it is evident that if C is the only parameter affected by pressure, α will also be affected by pressure to an amount equal to approximately one half the pressure effect on C. The pressure coefficient as α was therefore established as 0.05% per 500 pounds per square inch of pressure.

Laying Effect

Analysis of the Florida–Puerto Rico cable data indicated that the measured ocean bottom attenuation was less than the attenuation predicted from factory measurements. The differences were large enough to warrant study, and indicated that the measurements were in doubt, or that sea bottom conditions were not known accurately, or that some unexplained phenomenon was taking place.

In March 1955, approximately 22 mile of cable of the transatlantic design were laid in 300 fathoms of water off the coast of Spain in the Bay of Cadiz, and another equivalent length was laid in 2,300 fathoms off Casablanca. Precise measurements of attenuation were made in both cases, and it was established that a difference did, in fact, exist between measured values of attenuation at the ocean bottom and values predicted from factory measurements. It was further established that the difference in 2,300 fathoms was about twice that in 300 fathoms. The measured differences are shown in Fig. 14.

It was established during these trials that the difference increased slightly with time. Measurements made on the cable in 300 fathoms immediately, 18 hours, 48 hours, and 86 hours after laying indicated that measurable changes in attenuation were taking place. However, the change between 48 and 86 hours was so small, it was concluded that only very small changes would occur in a moderate interval of time. The tests also indicated that the attenuation of the two lengths of cable decreased somewhat during loading of the cable ship.

The total difference between the attenuation at the ocean bottom and the values predicted from factory measurements, taking the temperature and pressure coefficients into account, was designated "laying effect." Various theories, such as the consolidation of the central conductor, consolidation of return structure, and changes in the dielectric material, have been advanced to explain these differences. Each of these has been under study, but at the time of writing, no conclusive explanation has been established.

The shape of the "laying effect" versus frequency characteristic was such that the adjustment of repeater section lengths in conjunction with several fixed equalizers, which had approximately a 4-db loss at 160 kc and 0.6 db loss at 100 kc, would provide a good system characteristic. The matter of equalization is covered in greater detail in the article[4] on the over-all system. The magnitude of the laying effect observed during the laying of the two transatlantic cables substantiated the trial results.

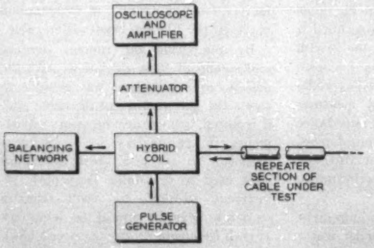

OSCILLOSCOPE AND AMPLIFIER
ATTENUATOR
BALANCING NETWORK
HYBRID COIL
PULSE GENERATOR
REPEATER SECTION OF CABLE UNDER TEST

Fig. 15. Simplified block diagram of pulse echo set for measurement of terminal impedance and internal impedance irregularities

Pulse Echo Measurements

Process controls, such as the use of a capacitance monitor and the jointing of the core in manufacturing sequence, provided the means for controlling the magnitude of reflections due to impedance mismatches. However, to ensure that the final product met these requirements, measurements of terminal impedance and internal irregularities were made with the use of pulse equipment.

A block schematic of the circuit of the echo set is shown in Fig. 15. For the submarine cable tests, a 1.5-microsecond raised cosine pulse was used, and the impedance of the balancing network was calibrated at 165 kc. The 165-kc impedance of the repeater sections was maintained well within a range of 54.8± 1 ohm. Internal irregularities at the point of irregularity were maintained at least 50 db below the magnitude of the measuring pulse. The requirement was 45 db.

References

1. Transatlantic Telephone Cable System—Planning and Over-all Performance, E. T. Mottram, R. J. Halsey, J. W. Emling, R. G. Griffith. Page 842 of this issue.

2. A Submarine Telephone Cable with Submerged Repeaters, J. J. Gilbert. *Bell System Technical Journal*, New York, N. Y., vol. 30, Jan. 1951, p. 65.

3. Air Force Submarine Cable System, P. T. Haury, L. M. Ilgenfritz. *Bell Laboratories Record*, New York, N. Y., Sept. 1956.

4. System Design for the North Atlantic Link, H. A. Lewis, R. S. Tucker, G. H. Lovell, J. M. Fraser. Page 850 of this issue.

System Design for the Newfoundland–Nova Scotia Link

R. J. HALSEY J. F. BAMPTON

Synopsis: The design and engineering of the section of the transatlantic cable system between Newfoundland and Nova Scotia was the responsibility of the British Post Office. Since the transmission objectives for this link were agreed upon in relation to the over-all objectives, this paper was written to show how these were translated into system and equipment design, and to demonstrate how the objectives were realized.

UNDER the terms of the agreement,[1] it was the responsibility of the British Post Office to design and engineer the section of the transatlantic cable system between Newfoundland and Nova Scotia. Like other parts of the system, all specifications were to be agreed upon between the Post Office and AT&T (American Telephone and Telegraph Company), but as both the British and the American types of submerged repeater had been carefully studied and generally approved by the other party prior to the agreement, the basic pattern of the system was clear from the beginning.

The service and transmission objectives for the over-all connections of London to New York and London to Montreal were agreed[2] in early joint technical discussions in New York and Montreal, and the agreed total impairments were divided appropriately between the various sections. In this way, the transmission objectives for the Newfoundland–Nova Scotia link were established.

Route

The choice of Clarenville as the junction point of the two submarine sections of the transatlantic system was determined primarily in relation to the Atlantic crossing and the desire to follow a transatlantic route to the north of existing telegraph cables.[3] There were a number of possibilities for the route between Clarenville and the east coast of Cape Breton Island, the most easterly point which could be reached reliably by the radio-relay system through the Maritime Provinces of Canada. One possibility, which had been considered earlier, was to cross Newfoundland by a radio-relay system and to use a submarine-cable link across Cabot Strait only. The final decision to build a cable system between Clarenville and Sydney Mines raised a number of problems in respect of the route to be followed which were concerned primarily with potential hazards to the cable brought about by:

1. The existence of very extensive trawling grounds on the Newfoundland Banks.
2. The location of considerable numbers of telegraph cables in the vicinity.
3. Grounding icebergs.

The route finally selected after thorough on-the-spot investigations[3–5] is satisfactory with respect to all these hazards, since it involves no cable crossings and is inshore of the main fishing grounds. This route is shown in Fig. 1. The straight-line diagram of the route is shown in Fig. 2. The total cable length is 326 nautical miles, of which 54.8 nautical miles are between Clarenville and Terrenceville, Newfoundland, where the cable finally enters the sea. The maximum depth of the water involved is about 260 fathoms.

Cable

Choice of Design

Since 1930, when the Key West–Havana no. 4 cable was constructed,[6] it has been usual to extrude the insulation of coaxial submarine cables to a diameter of about 0.62 inch, and most of the cables in the waters around the British Isles are of this size. The experience of the British Post Office with submerged repeaters[7] in its home waters, dating from 1944 when the first repeater was laid between Anglesey and the Isle of Man,[8] has therefore been mainly with 0.62 inch cables, first with paragutta as a dielectric and later with polyethylene. Most of these cables were originally operated without repeaters, and the 60-circuit both-way repeaters which are now installed on the routes were designed to match their characteristics.

In planning a new system, the size of cable will be determined by one of the following considerations:

1. Minimum annual charges for the desired number of circuits.
2. Terminal voltage required to feed the requisite number of repeaters.
3. Maximum number of repeaters or minimum repeater spacing which is considered permissible.
4. Maximum or minimum size of cable which can be safely handled by the laying gear in the cable ship.

When the 36-circuit system between Aberdeen, Scotland, and Bergen, Norway, was planned in 1952, the route length of 300 nautical miles greatly exceeded

---

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957.

R. J. Halsey and J. F. Bampton are with the British Post Office, London, England.

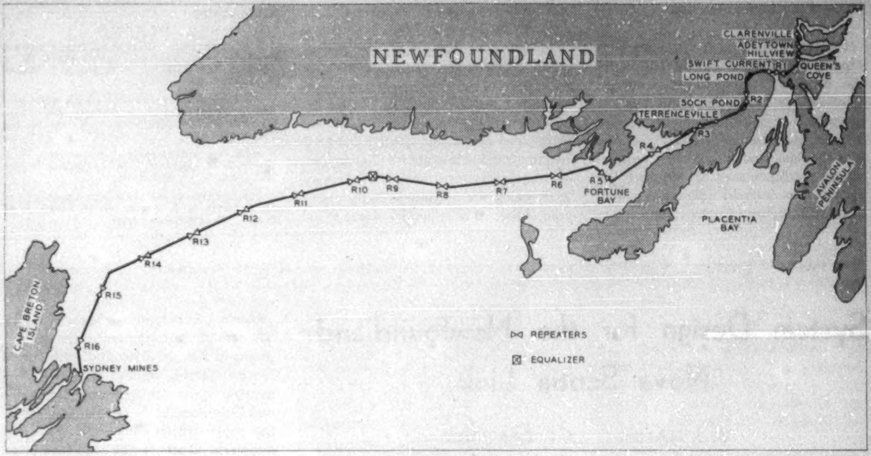

NEWFOUNDLAND
CLARENVILLE
ADEYTOWN
HILLVIEW
SWIFT CURRENT
R1
QUEEN'S COVE
LONG POND
SOCK POND
R2
TERRENCEVILLE
R3
R4
R5
FORTUNE BAY
R6
R7
R8
R9
R10
R11
R12
R13
R14
R15
R16
AVALON PENINSULA
PLACENTIA BAY
CAPE BRETON ISLAND
SYDNEY MINES
REPEATERS
EQUALIZER

Fig. 1. Map of route

that of any other submarine telephone system, and it was decided to use a core diameter of 0.935 inch, first, to keep the number of repeaters as low as seven, and second, because the system was intended as a prototype of a possible Atlantic cable. The cable dielectric is polyethylene grade *2* with 5% (per cent) of polyisobutylene.

For the Clarenville–Sydney Mines link it proved possible to design for minimum annual charges. With increasing experience and confidence in submerged repeaters, it was no longer considered necesssary to restrict the number of repeaters as for Aberdeen to Bergen, and the terminal-voltage requirements were reasonable. At the current prices of cable and repeaters in Great Britain, the optimum core diameter for 60 both-way circuits is about 0.55 inch, but the increased charge incurred by using 0.62-inch cable is less than 5% (0.62-inch core is optimum for 120 both-way circuits). To facilitate the manufacture and the provision of spare cable, it was therefore logical to adopt the same design as that proposed for the Atlantic crossing. This is described in other papers.[3,9]

After investigating various possible types of cable for the overland section in Newfoundland, it was decided to use a design essentially the same as the main cable, but with additional screening against external interference.[4] For the outer conductor and its copper-binding tape, the construction is identical with that of the main cable except that the compounded-cotton tape is overlapped. This construction is shown in Fig. 3. Outside this are five layers of soft-iron tapes each 0.006 inch thick, the innermost being longitudinal and the others having alternate right- and left-hand lays at 45 degrees to the axis of the cable. After another layer of compounded-cotton tape, there is extruded a polyethylene sheath 0.080 inch thick, and the whole is jute-served and wire-armored. As a check on the efficiency of the screening, the maximum sheath-transfer impedance at 20 to 100 kc was specified as 0.005 ohm per 1,000 yards.

It was thus possible to treat the entire link from Clarenville to Sydney Mines as a uniform whole, using the same type of repeater on the land as in the sea. A small hut at Terrenceville contains passive networks only.

Attenuation Characteristics

When the system was designed, precision measurements of cable attenuation were not available. The design of the Oban-Clarenville link was based on laboratory measurements on earlier 0.62-inch cable of a similar type, but the available data applied only to frequencies up to about 180 kc, whereas the Clarenville–Sydney Mines link was to operate at frequencies up to 552 kc. Extensive extrapolation was therefore involved. As soon as the first production lengths of cable became available in February 1955, laying trials were conducted off Gibraltar. It was found that there were serious changes of attenuation on laying, over and above those directly attributable to temperature and pressure effects, and that the assumed characteristics were inaccurate. Although the attenuation in the factory tanks had been in reasonable agreement with that of the earlier cable, there were changes on transfer to the ship's tanks and again on laying, amounting in all to a reduction of about 1.5% at 180 kc. This would have been comparatively unimportant

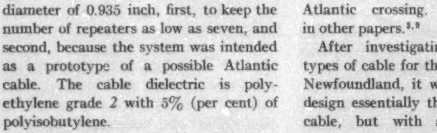

ADEYTOWN
QUEENS COVE S. W. ARM
RANDOM SOUND
PLACENTIA BAY
HILLVIEW
TERRENCEVILLE
R2
R1
CLARENVILLE
12.31
20.02
22.52
SYDNEY MINES
R16
R10
E
R9
R3
TERRENCEVILLE
11.68
6.094
5.707
2.932

Fig. 2. Straight-line diagram of route. All distances are in nautical miles, and spacing R3 to R9, and R10 to R16 is 20.4 nautical miles

R—Repeater
E—Equalizer

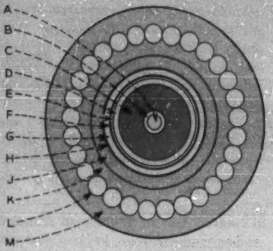


Fig. 3. Cross section of cable across Newfoundland, showing make-up

A—Center conductor, 0.1318-inch copper diameter
B—Three 0.0145-inch copper surround tapes
C—Polyethylene to 0.620-inch diameter
D—Six 0.016-inch copper return tapes
E—0.003-inch overlapped copper teredo tape
F—Impregnated-cotton tape
G—Five iron-screen tapes
H—Impregnated-cotton tape overlapped
J—Polyethylene sheath to 1.02-inch diameter
K—Inner serving of tanned jute yarn
L—Armor wire 29 × 0.128-inch diameter
M—Outer serving of tarred jute yarn

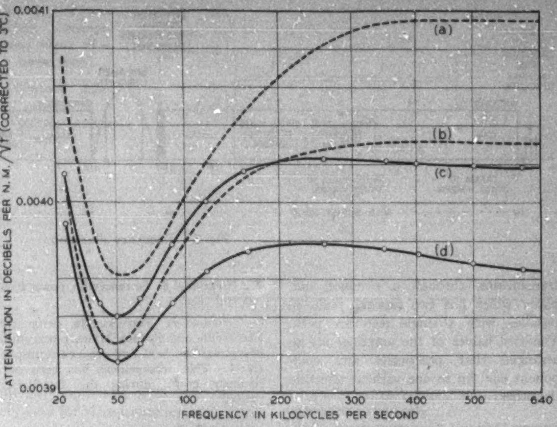


Fig. 4. Cable attenuation characteristics, Skye trials

(a)—Characteristic originally assumed
(b)—Characteristic measured in factory tank (flooded)
(c)—Characteristic measured in ship's tank (flooded)
(d)—Characteristic measured after laying

had the discrepancy been of "cable shape," i.e., the same fraction of the cable attenuation at all frequencies and therefore exactly compensated by a length adjustment of the repeater sections. As this was not so, and as the cable-equalizing networks in the repeaters were settled by this time, it was clear that precise information must be obtained so that suitable additional equalizers could be provided for insertion in the cable on laying. There are a number of factors which may lead to small changes of attenuation on laying, but most of these tend to increase the losses. The primary reason for the observed changes appears to be contact variations between the various elements of the inner and outer conductors, i.e., the wire and three helical tapes forming the center conductor and the six helical tapes forming the return conductor. These contact resistances tend to change with handling and, as a result of a slight degree of "bird caging" when coiled, it seems that the attenuation decreases as the coiling radius increases, and vice versa. Also, the effect of sea pressure is to consolidate the conductors and thereby further reduce the attenuation. This effect appears to continue on a diminishing basis for a long time after laying.

To obtain reliable data for the Clarenville–Sydney Mines link, 10 nautical miles of cable with *A*-type armor were laid off the Isle of Skye at about the mean depth of the system, 120 fathoms. The attenuations, coiled and laid, are shown in Fig. 4, with due allowance made for temperature and pressure. The ordinates, attenuation versus frequency, are such that the value should be approximately constant at high frequencies.

In making a final determination of the cutting lengths for the repeater sections, it was assumed that the factory measurements of attenuation would be reduced by 1.42% at 552 kc on laying, that the temperature coefficient of attenuation would be +0.16% per degree centigrade, and that the true pressure coefficient of attenuation was negligible at the depths involved.

Description of System

Circuit Provision and Frequency Allocation

It was originally thought that a design similar to that of the Aberdeen-Bergen system would be suitable for the Clarenville–Sydney Mines route in that it would provide 36 more circuits than the long section across the Atlantic. This potential excess capacity, which was required for circuits between Newfoundland and the Canadian mainland, disappeared when it was found that 36 circuits could, in fact, be provided over the longer link. The Aberdeen-Bergen design was therefore modified to provide a complete supergroup of 60 circuits, the same capacity as the earlier British projects.[7] The system thus requires broad-band transmission of 240 kc in each direction.

In the earlier British projects, the transmitted frequency bands were 24 to 264 and 312 to 552 kc, but for the present purpose the lower band is dropped by 4 kc to 20 to 260 kc, so that the lowest frequency is the same as on the Atlantic cables. This enables common frequency-generating equipment to be used at Clarenville for the two links and minimizes crosstalk problems. The main transmission bands and the allocation of the five 12-circuit groups are shown in Fig. 5, together with the ancillary channels. The facilities provided are discussed later.

Submerged Repeaters

The submerged repeaters which were used are fully described elsewhere,[10] and it will suffice to note here that they are rigid units, approximately cylindrical in shape, 9 feet long, and have a $10\frac{1}{2}$-inch maximum diameter. They are capable of withstanding the full laying pressure in deep water, although this is of little importance in the present application.

They are arranged for both-way

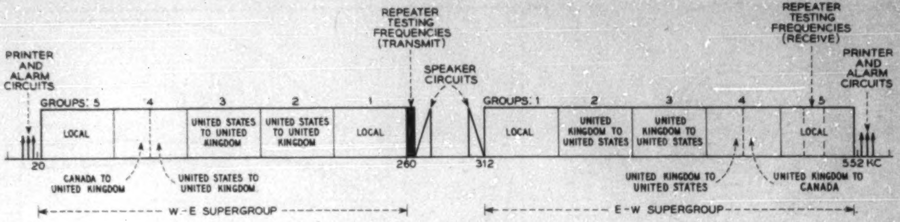

REPEATER TESTING FREQUENCIES (TRANSMIT)
SPEAKER CIRCUITS
REPEATER TESTING FREQUENCIES (RECEIVE)
PRINTER AND ALARM CIRCUITS
PRINTER AND ALARM CIRCUITS
GROUPS: 5
4
3
2
1
LOCAL
UNITED STATES TO UNITED KINGDOM
UNITED STATES TO UNITED KINGDOM
LOCAL
GROUPS: 1
2
3
4
5
LOCAL
UNITED KINGDOM TO UNITED STATES
UNITED KINGDOM TO UNITED STATES
LOCAL
20
260
312
552 KC
CANADA TO UNITED KINGDOM
UNITED STATES TO UNITED KINGDOM
UNITED KINGDOM TO UNITED STATES
UNITED KINGDOM TO CANADA
W-E SUPERGROUP
E-W SUPERGROUP

Fig. 5. Frequency allocation

transmission through a common amplifier which has two forward paths in parallel, with a single feedback path. The two halves of the amplifier are so arranged that practically any component can fail in one without affecting the other.

Power-Feeding Arrangements

The submerged repeaters are energized by constant-current d-c supplies between the center conductor and ground. The power units at the two ends are in series aiding, and the repeater power circuits are in series with the center conductor, i.e., without ground connections, as in Fig. 6. This is the only arrangement by which it is possible to control the supply accurately at every repeater, the insulation resistance of cable and repeaters being sufficiently great that the current in the center conductor is virtually the same at all points. The constant-current feature of the supply ensures that repeaters cannot be overrun in the event of a ground fault on the system.

With this arrangement the link requires a total supply voltage of about 2,300 volts. The power-feeding equipment[12] at each terminal station is designed to feed a constant current of 316 milliamperes at this voltage, and it is permissible to energize the system from one end only, if necessary. The repeater capacitors which are the limiting factors in respect to line voltage are rated very conservatively at 2,500 volts, so that a single-ended supply of 2,300 volts, with the possibility of superimposed ground-potential differences, is near the desirable maximum. The two terminal power units are therefore designed to operate in series and to share the voltage.

With access to the cable provided at Terrenceville, it is possible to operate the power system on the following bases:

1. No ground at Terrenceville, power from both ends on a master-and-slave basis to ensure that the constant-current units do not build up an excessive voltage. This is the normal arrangement.
2. No ground at Terrenceville, power from one end only.
3. Ground at Terrenceville, with the Clarenville and Sydney Mines power units energizing the land and sea cables respectively. This arrangement has been particularly useful during the installation period.

The presence of high voltages on the cable constitutes a potential danger to personnel, therefore special precautions are taken in the design of the equipment in which the cable terminates and in which high voltages exist or may exist.

The ground connections for the power circuits at the two ends are via special ground cables, and ground plates are located about half a mile from the main cable. Metering arrangements are provided to check that the current does in fact take this path. These measures ensure that the current returning via the cable armor is never sufficient to cause serious corrosion.

Arrangement of Terminal Equipment

Fig. 7 shows the arrangement of the terminal equipment. In accordance with an early agreement precisely defining the various sections of the project, the link is considered to terminate at the group distribution frames at Clarenville and Sydney Mines, i.e., at the 60–108-kc interconnection points.

In addition to the cable-terminating and power-feeding equipments, *A* and *B*, the following are provided at the terminals:

1. Submarine-cable terminal equipment, *C*, consisting of repeaters to amplify the signals transmitted to and received from the cable, and equalizers and frequency-translating equipment to convert the line frequencies to basic supergroup frequencies, 312 to 552 kc.
2. Group-translating (group-bank) equipment, *D*, to convert the basic supergroup to five separate basic groups, 60 to 108 kc, and vice versa.
3. Equipment for the location of cable and repeater faults, *E* and *F*.
4. Speaker and printer circuit equipments, *G* and *H*, to provide two reduced-bandwidth telephone circuits, two telegraph circuits, and one alarm circuit for maintenance purposes. It is clear that such circuits should be substantially independent of the main transmission equipment.

Two principles were agreed upon very early in the planning; first, that the engineering of the various links should be integrated as far as possible, and second, that the items of equipment at each station should be provided by the party best in the position to do so. The result was that:

1. Items of standard equipment were provided by AT&T at Clarenville and Sydney Mines, and by the Post Office at Oban, thus simplifying maintenance and repair problems.
2. Basic power plant and the carrier supplies for supergroup and group translation were provided by AT&T for both terminal equipments at Clarenville.
3. Terminal equipment, special to the Clarenville–Sydney Mines link, was provided by the Post Office.

In view of the importance of the link, the power-feeding and transmission equipment are completely duplicated.

The submarine-cable terminal equipment is arranged to transmit the basic supergroup, directly over the cable in the east-to-west direction. In the west-to-east direction the supergroup is translated to the range 20 to 260 kc, using a 572-kc carrier.

Design of Transmission System

Performance Requirements

The agreed transmission objects for the Clarenville–Sydney Mines link were as follows:

Variation of Transmission Loss. The variation in the transmission loss of each group should have a standard deviation not greater than 0.5 db (decibel). This implies that the variation from nominal should not exceed 1.3 db for more than one per cent of the time.

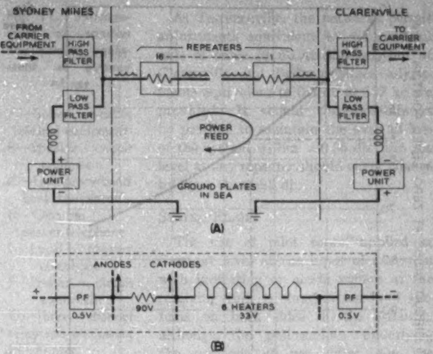
Attenuation/Frequency Characteristics. Only the over-all characteristics of the individual circuits were precisely specified, the limits being the CCIF (International Consultative Committee on Telephony) limits for a 2,500-km (kilometer) circuit and the target one half of this. With this objective in view, the group characteristics in each link must clearly be as uniform as is practicable.

Circuit Noise. The total noise contributed by the link to each channel in the busy hour, i.e., including intermodulation noise, should have an rms value not exceeding +28 dba (decibels adjusted),* which corresponds to −56 dbm (decibels referred to one milliwatt), at a point of zero relative level.

Crosstalk. The minimum equal-level crosstalk attenuation should be 61 db for all sources of potentially intelligible crosstalk; this was accepted as a target for both near- and distant-end crosstalk. Although go-to-return crosstalk is not important for telephony (it appears as sidetone) and a limit of 40 db is satisfactory even for voice-frequency telegraphy, it assumes great importance for both-way music transmission. Also, it was desired to be nonrestrictive of future usage.

Fig. 6. Power-feeding arrangements, where PF = power filter

A—General schematic
B—Repeater power circuit


SYDNEY MINES
FROM CARRIER EQUIPMENT
HIGH PASS FILTER
LOW PASS FILTER
POWER UNIT
+
−
REPEATERS
16
1
POWER FEED
GROUND PLATES IN SEA
CLARENVILLE
TO CARRIER EQUIPMENT
HIGH PASS FILTER
LOW PASS FILTER
POWER UNIT
−
+
(A)
ANODES
CATHODES
+
PF
0.5V
90V
6 HEATERS
33V
PF
0.5V
−
(B)

Assessment of Requirements

The design of the high-frequency path to meet the agreed requirements involves consideration of the following:

1. Noise, including fluctuation (resistance and tube) noise and intermodulation.
2. Wide-band frequency characteristics, including the effects of the directional filters at the terminal and in the repeaters.
3. Variations of 1 and 2 with respect to temperature and aging.

The noise requirement is by far the most important factor in the design of the line system.

The choice of route and cable having been made, the total loss was known, and it was necessary to determine the minimum number of repeaters to compensate for this loss and to meet the noise requirement with adequate margin for inaccurate estimates of capable attenuation after laying, temperature variations, aging, and repairs. An attempt to achieve the necessary gain with too few

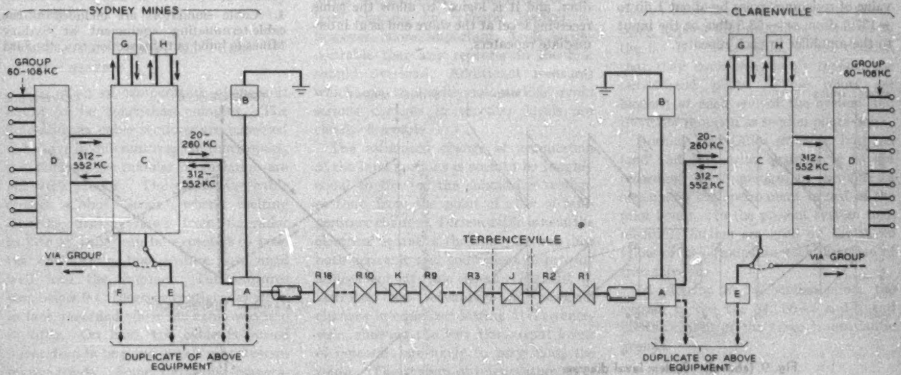

SYDNEY MINES
GROUP 60–108 KC
G
H
D
C
312–552 KC
20–260 KC
312–552 KC
B
VIA GROUP
F
E
A
DUPLICATE OF ABOVE EQUIPMENT
TERRENCEVILLE
R16
R10
K
R9
R3
J
R2
R1
CLARENVILLE
GROUP 60–108 KC
G
H
B
20–260 KC
312–552 KC
C
312–552 KC
D
VIA GROUP
A
E
DUPLICATE OF ABOVE EQUIPMENT

Fig. 7. Arrangement of terminal equipment and repeaters

A—Cable-terminating equipment
B—D-c power-feeding equipment
C—Submarine-cable terminal equipment
D—Group-translating equipment
E—Pulse-test equipment
F—Loop-gain test equipment
G—Speaker circuit equipment
H—Printer circuit equipment
J—Intermediate cable-terminating equipment, including equalizers
K—Submerged equalizer
R—Submerged repeater

---

* This refers to the reading on a Bell System *2B* noise meter (F1A weighting network). The noise level, dba, is relative to a 1.0-kc tone at −85 dbm. In Europe, noise is measured on a CCIF psophometer (1951 weighting network), which is calibrated in millivolts across 600 ohms. This is commonly converted to picowatts (pw). The white noise equivalence of the two instruments is given by dba = 10 $\log_{10}$ pw − 6 = dbm + 84. The agreed limit of +28 dba is therefore equivalent to 2,513 pw (1.23 millivolts or −56 dbm). The corresponding CCIF requirement at 4.0 pw per km would be 2,400 pw, this value not to be exceeded for more than one per cent of the time.

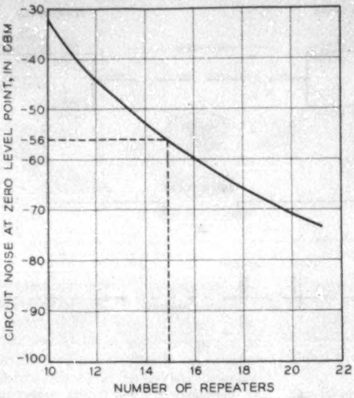


Fig. 8. Variation of circuit noise with number of repeaters

repeaters would result in excessive noise.

Design of the amplifiers in the British repeaters is such that, with both forward paths in operation, the overload point is about +24 dbm, and with a loading of 60 channels in each direction, this permits planning levels of about −4 dbm at the amplifier output after allowing reasonable margins for errors and variations.[10] Previous experience shows that, at such output levels, intermodulation noise can be neglected and the full noise allowance allotted to fluctuation noise. The effect of tube noise is to increase the weighted value of resistance noise by about 1 db to −137.5 dbm, or −53.5 dba, at the input to the amplifier in each repeater.

At the highest transmitted frequency the equalizers, power filters, and directional equipment introduce losses of about 1 db and 4 db at the input and output of the amplifier respectively; these losses must, effectively, be added to the loss in the cable.

The two other items of information which are necessary before the repeater system can be planned are the permissible transmitting and receiving levels at the shore stations. The transmitting equipment, provided at Clarenville can be operated at channel levels up to +20 dbm, and it is logical to allow the same receiving level at the shore end as at intermediate repeaters.

On this basis it is possible to construct a curve, Fig. 8, relating the total circuit noise to the number of intermediate repeaters, and it is seen that the minimum number is 15, each of which must have an over-all gain of 59 db (amplifier gain, 64 db) at 552 kc. The actual provision is for 16 repeaters, each having a gain of 60 db at 552 kc, the additional gain being absorbed in fixed and adjustable networks at points along the route, as indicated in the following section.

Level Diagram

The actual level diagram (planning levels are shown in Fig. 9) differs somewhat from that which can be deduced directly from the preceding section because of the following considerations:

1. The location of the first repeater from Clarenville, i.e., on land, was dictated by topography and the desire to locate both it and the second repeater in ponds. Thus, the transmitting level at Clarenville is substantially lower than the permissible maximum.
2. There are equalizing networks at Terrenceville and facilities for their adjustment to compensate for temperature variations.
3. Because of the difference between the actual cable attenuation and that for which the repeaters were planned (see the section "Attenuation Characteristics"), it was necessary to include an equalizer unit in the sea, midway between Terrenceville and Sydney Mines. Loss equivalent to 9 nautical miles of cable was also introduced at this point to ensure that repeater no. 16 would be sufficiently far from the shore at Sydney Mines.
4. Cable simulators are included in the cable-terminating equipment at Sydney Mines to build out this section to a standard

Vol.
Pt. 1
1956

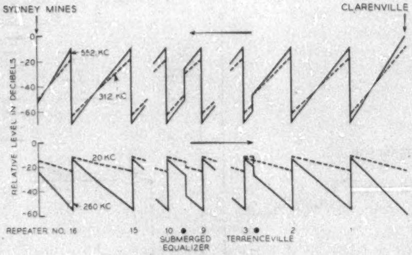


Fig. 9 (above). System level diagram

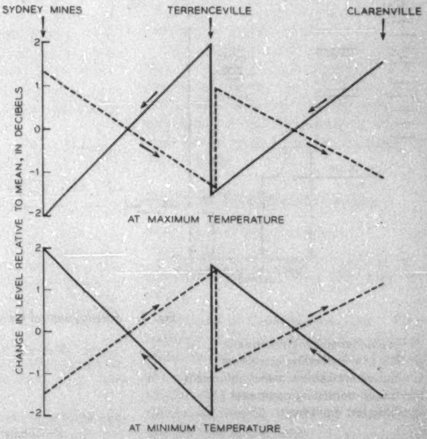


Fig. 10 (right). Deviation from mean of transmission levels with optimum adjustments of equalizers at Sydney Mines, Terrenceville, and Clarenville. Maximum deviation in two directions occurs at these frequencies

---West to east at 260 kc
——East to west at 553 kc

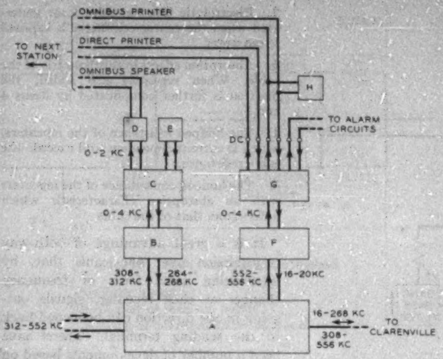


Fig. 11. Arrangement of speaker and printer equipment at Sydney Mines

A—Submarine-cable terminal equipment
B—Speaker - circuit equipment
C—Emergency-band bank equipment
D—Omnibus-speaker telephone
E—Local - speaker telephone
F—Printer - circuit equipment
G—Three - channel telegraph equipment
H—Printer

repeater section. The actual cable length was, of course, unknown until the cable was complete.

Taking into account the existence of the intermediate networks, the repeater spacing is such that, when both land- and submarine-cable sections are at mean temperature, the compensation is as accurate as possible. In general, the highest frequency is of greatest importance in this respect. Since the low-frequency channels experience less attenuation than the high-frequency channels, it is permissible to transmit them at a somewhat lower level, thereby increasing the load capacity of the amplifiers, which is available to the high-frequency channels.

TEMPERATURE EFFECTS AND THEIR COMPENSATION

The effect of temperature changes is likely to be somewhat complex. The land-and-sea cable sections are expected to behave in different ways in this respect, but data on the manner of variation are not very precise. The submarine cable crosses Cabot Strait, where melting icebergs, drifting down from Labrador as late as June, can be expected to keep the sea-bottom temperature low until well into the summer. Temperatures just below 0 C (degrees centigrade) were, in fact, recorded when the cable was laid in May. On land, the cable is buried 3 feet deep in bog and rock, and traverses many ponds. Some data on temperatures during similar conditions in other parts of the world were available.

For planning purposes, it was clear that the assumptions made would have to be somewhat pessimistic, and the assumed ranges of temperature, with the corresponding changes of attenuation at 552 kc, were:

Sea section............2.3±3 C; ±4 db
Land section...........7.5±10 C; ±3 db

A possible method of circuit adjustment for temperature changes is to increase the gains equally at the sending and receiving terminals, as the temperature rises, and to reduce them equally as it falls. Under such conditions, the effect of temperature variations on resistance noise is not very important; the levels at repeaters near the center of the route remain substantially constant, and the increase in noise from the repeaters whose operating levels are reduced is partly compensated by the reduction in noise from those whose levels are increased. The effect of the level changes on repeater loading is, however, more important, as it is undesirable that any repeater in the link should overload. Additional measures which can be readily adopted to avoid serious changes in repeater levels are clearly desirable.

The estimated change of attenuation of the land sections is seen to be roughly equal to that of the submarine section, so that, from the point of view of temperature changes, Terrenceville is near the electrical center of the link. It was thus both desirable and convenient to provide adjustment at this point. Fig. 10 illustrates the advantage of seasonal changes in equalizer setting at Terrenceville, showing the way the output levels of repeaters are likely to vary along the route. The system of temperature compensation which was adopted therefore involves adjustable networks at both ends of the system and at Terrenceville. All the networks are cable simulators, thus, the process of temperature compensation consists, effectively, in adding cable when the temperature falls and removing it when the temperature rises.

At Terrenceville, the networks permit adjustments equivalent to ±1 nautical mile of cable (3 db at 552 kc), but at Clarenville and Sydney Mines adjustments equivalent to 0.5 db at 552 kc are provided. It should therefore always be possible to maintain the over-all loss of the system within ±0.25 db, and the level at any repeater should never change by more than ±2 db.

SYSTEM PILOTS

The use of pilot tones, applied at constant level at the input of a system with indicating or alarm meters at the receiving end, is standard on land systems on both sides of the Atlantic, although the philosophies underlying the methods of use differ. On the submarine cables round the British Isles, with or without submerged repeaters, pilot tones are used to indicate the attenuation of the transmission path. These pilots are normally located just outside the main transmission bands in each direction. In the Clarenville–Sydney Mines system, the frequency bands just outside the main transmission bands are occupied by telephone speaker and teleprinter circuits and by monitoring frequencies associated with the repeaters (see Fig. 5). This prevents the use of out-of-band pilots.

Fortunately, the standard Bell System group equipment is designed to apply 92-kc pilots to each group and to measure the corresponding received level. Although these are essentially group pilots, being applied and measured at points in the 60- to 108-kc band, it was decided that they could reasonably replace the out-of-band pilots. These pilots are blocked at each end of the system and therefore function as section pilots only.

Normal Post Office practice, both on land and submarine systems, is to use recording level meters to provide a continuous and permanent record of the pilot levels. In the present system such recording meters are used on the 92-kc pilots of two groups, in each direction of transmission.

In addition to the section pilots, the system carries the 84.080-kc end-to-end pilots in each of the three transatlantic groups.

Maintenance Facilities

SPEAKER AND PRINTER CIRCUITS

It was part of the planning of the transatlantic system that two low-grade telephone (speaker) and two telegraph (printer) circuits should be provided over the submarine cables, outside the

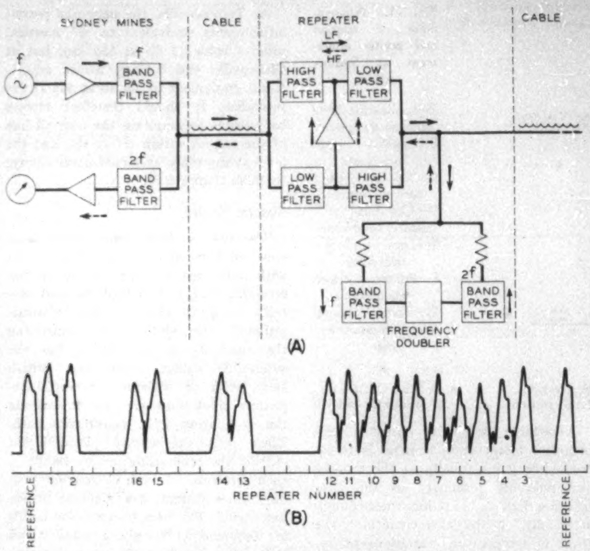


Fig. 12. Fault location, loop-gain method

A—Block schematic, frequency f is in the band 260 to 264 kc
B—Diagram of display

main transmission bands, and that the speaker circuits, in particular, should be reasonably independent of the main terminal equipment. One speaker circuit is needed for local communication between the terminals of each section, the other is needed to form part of an omnibus circuit which connects the principal stations on the route, including Montreal. The arrangement for teleprinter communication was that one channel should be an over-all all-station omnibus printer, while the other should be a direct London–New York printer.

Independent frequency-translating equipment is provided to connect the speaker and printer bands, each 4 kc, to the line. The carrier frequencies required for the speaker are provided by independent oven-controlled crystal oscillators, but, for the printer, the independent generation of high-stability 572-kc carriers was not considered to be justified, and the main station supplies are used.

Two half-bandwidth telephone circuits are provided in the 4-kc speaker bands by the use of standard AT&T band-splitting equipment (EB banks). Signalling and telephone equipment are provided to give the required omnibus facilities on one circuit and local-calling facilities on the other. The arrangement of the speaker and printer equipment at Sydney Mines is shown in Fig. 11.

In the telegraph band, a third channel transmits an alarm to the remote terminal when the 92-kc pilots, coming in from that terminal, fail simultaneously.

FAULT LOCATION

The speedy and accurate location of faults in repeatered cables is of very great importance, because of the number of circuits involved and the difficulty and cost of repairs. The standard d-c methods which have been applied in the past to long telegraph cables are, of course, available. The application of these methods is, however, recognized as being rather more in the nature of an art than a science and usually requires an intimate knowledge of the behavior and peculiarities of the particular cable concerned. While the problem appears at first to be simple, it is complicated by:

1. The presence of ground-potential differences along the cable, sometimes amounting to hundreds of volts. These vary with time.
2. Electrolytic electromotive force generated when the center conductor is exposed to sea water.
3. Absorption effects in the dielectric of the cable. When repeaters are added, the position is further complicated by items 4 and 5.
4. The lumped resistance of the repeaters, which is current dependent and exceeds the cable resistance.
5. The lumped capacitance of the repeaters with an absorption characteristic which differs from that of the cable.

It is a great advantage of both-way transmission over one cable that, by introducing some form of frequency changer at each repeater, signals outgoing in one direction can be looped back to the sending terminal. There have been a number of developments based on this principle and, in the Clarenville–Sydney Mines link, two methods are available for use. Of these, the so-called "loop-gain" method uses steady tones and depends on selective frequency measurements to discriminate between repeaters; the second is a pulse method in which repeaters are identified on the basis of loop transmission time.

The use of these methods during fault conditions depends on the possibility of keeping the repeaters energized. Work is in progress to develop methods of fault location which are of general application and do not depend on the activity of the repeaters, but these are outside the scope of this paper.

Loop-Gain Method

In the loop-gain method, the frequency changer in the repeater takes the form of a frequency doubler, and each repeater is identified uniquely by one of a group of frequencies spaced at 120 cycles and located immediately above the lower main transmission band in the frequency range of 260 to 264 kc. Since the frequency changing is in an upward sense, the measuring terminal is Sydney Mines, which transmits the lower band. On the Clarenville side of the directional filters, in each repeater is connected, via series resistors, a crystal filter accepting the test frequency appropriate to the repeater; see Fig. 12(A). This frequency is doubled, filtered, and returned to the repeater at the same point at which the original frequency is selected. From this point it passes via the high-pass directional filters through the amplifier and back to Sydney Mines, where the level is measured on a transmission-measuring set. Information obtained in this way on each repeater can be compared at any time with that obtained when the system was installed, and any

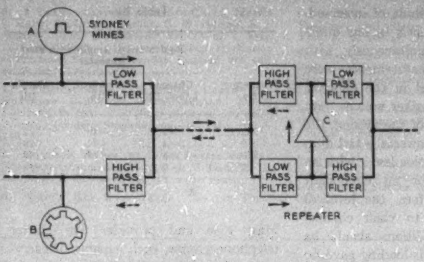


Fig. 13. Fault location, pulse method

A—Pulse generator
B—Display of received pulses
C—Point of intermodulation, output of amplifier

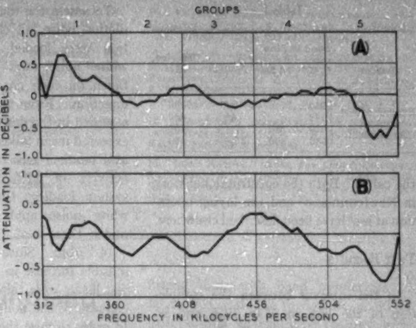


Fig. 14. Attenuation versus frequency characteristics of supergroup

A—Sydney Mines to Clarenville
B—Clarenville to Sydney Mines

gain variations can be localized. The test equipment provided has the additional facility of an automatic sweep of the test frequency at 4 cycles and has a display of the returned-signal levels on a cathode-ray tube as shown in Fig. 12(B).

Although the transmitted signals lie outside the band of the west-east supergroup, the received signals, 520 to 528 kc, lie within the band of the east-west supergroup, and two channels must be removed from traffic to carry out the tests. These channels are in a "local" group.

Pulse Method

As applied to the present system, the pulse method uses the overload characteristic of the amplifier to effect the frequency change in the repeater. At Sydney Mines, a continuous train of single-frequency pulses is applied in the lower transmission band, so that either the second or third harmonic is returned in the upper band, as shown in Fig. 13. At Clarenville, 2-frequency pulses are applied in the upper band, so that either a second- or third-order difference product is returned in the lower band. The pulse length is 0.15 millisecond, and the frequencies used are given in Table I. At Sydney Mines, the signals can be sent and received either on the line itself or via the group equipment. In the latter case, only one group need be taken out of service. At Clarenville line, only measurements are provided for.

The primary display is on a cathode-ray tube with a circular time base, and any one returned pulse can be accurately compared with the reference pulse on a second tube with a linear time base. The pulse selected for such measurement is automatically blacked out on the primary display.

Usefulness of Loop-Gain and Pulse Methods

Both methods require that all the repeaters between the testing terminal and the fault can be energized. If the fault is in the cable, there is a very high probability that the center conductor will be exposed to the sea, in which case the power circuit can be maintained at least on one side of the fault, although it may be somewhat noisy. Because the system is short, it is permissible to energize the link fully from one end only. The condition can never arise, as it can in the Oban-Clarenville link, that the line current is limited by the maximum permissible terminal voltage.

The loop-gain test is concerned with the amplifiers in their linear regime and gives no indication of the overload point. For this, the pulse test must be used. On the other hand, the pulse test does not permit accurate measurement of levels, since the pulse level which reaches a particular repeater may be restricted by the overload of an earlier repeater in the chain. The pulse test is particularly useful in providing a check to see that both sides of each amplifier are in operation and to locate a fault of this type.

Each method, for its operation, depends on nonlinearity at a point within each repeater and can only identify a fault as lying between two such consecutive points in the link. It is therefore desirable that these points should be as close as possible to the terminals of the repeater to ensure that the faulty unit can be identified. In this respect, the loop-gain test has the advantage over the pulse test.

Execution of Work

Problems caused by the remoteness of the site were overcome without undue difficulty with the co-operation of the other parties concerned with the project, but this paper would be incomplete without a brief reference to the cable- and repeater-laying operations in Newfoundland and at sea.

The terrain and conditions in Newfoundland were quite unlike those with which the British Post Office normally has to contend, since they involved trenching and cabling through bog, rock, and ponds in country of which no detailed survey or maps were available. Maps were constructed from aerial survey, and alternative routes were explored on foot before a final choice was made. As much use as possible was made of water sections in the sea, river estuary, and ponds; some 22 miles were accounted for in this way, leaving about 41 miles to be trenched by machine or blasted. A contractor was engaged for this purpose and to lay the cable in the trench, but all jointing was done by the Post Office. The standards of conductor and core jointing were the same as those in the cable factories and on ship, with portable injection-moulding machines and X-ray equipment specially designed for handling over the bog. A single-pair cable was also laid in the main cable trench to provide speaker facilities between Clarenville and Terrenceville, which has no public telephone, with intermediate positions for use of the lineman. As a measure of protection against lightning strikes, two bare copper wires were buried about 12 inches apart and 6 inches above

Table I

Station	Send to Line		Product	Receive, Kc
	f_1, Kc	f_2, Kc		
Sydney Mines	216		$2f_1$	432
	144		$3f_1$	432
Clarenville	530	380	f_1-f_2	150
	530	340	$2f_2-f_1$	150

the cable. Both the constructional work in Newfoundland[4] and the laying operation at sea[5] have been described elsewhere.

Test Results

In the interval between the completion of the link in May 1956 and its incorporation in the transatlantic system, tests were carried out to establish its performance and day-to-day variations. An assessment of the annual variations has, of course, been impossible at present.

VARIATION OF TRANSMISSION LOSS

Close observation of the transmission loss of the 92-kc pilots on groups 1 and 5 leads to these tentative conclusions:

1. Over periods of one hour, the variations are not measurable, i.e., less than ±0.05 db.
2. Over periods of 24 hours, there are no systematic changes; apparently random changes of about 0.1 db are probably attributable to the measuring equipment.
3. Over a period of 8 weeks, July and August 1956, there was a systematic increase in loss of about 0.3 db. By means of the loop-gain equipment, it has been possible to deduce that most of this change has occurred in the land section.

The results indicate that the submarine-cable link has better day-to-day stability than the best testing equipment which it has been possible to provide. Many more data will clearly be necessary before the annual variations can be established definitely, but the present indications are that these variations will be less than those assumed in the design of the link.

ATTENUATION/FREQUENCY CHARACTERISTICS

The frequency characteristics of the supergroup in the two directions of transmission are shown in Fig. 14. It will be seen that in no transatlantic groups does the deviation from mean exceed ±0.35 db.

CIRCUIT NOISE AND CROSSTALK

Table II shows the noise level on channels 1 and 12 of each of the five groups measured without traffic on the system.

To assess the magnitude of intermodulation noise, all channels in one direction were loaded simultaneously with white noise, and measurements were taken on each channel in the opposite direction. From the talker volume data assumed in the design of the system, the expected mean talker power is −11.1 dbm at a point of zero relative level, with an activity of 25%. For an equivalent system loading, therefore, the level of white noise applied to each channel under these test conditions should be −14.1 dbm. Since this loading gave no sensible increase in the circuit noise, the test levels were raised until a reasonable increase in the noise level was obtained. In order to raise the channel noise to the specified maximum of 28 dba, it was necessary to raise the channel levels to about −1 dbm and −4 dbm in the lower and upper bands, respectively. These levels, some 13 db and 10 db above the assumed maximum loading of the system, give noise levels at least 26 db and 20 db above normal, and it is seen that adequate margins exist for variations and deterioration of the link.

Closely allied to the problem of intermodulation is the overload characteristic of the system. Table III shows the measured overload point of the link, expressed as an equivalent level at the output of the amplifier in the repeater nearest to the transmitting terminal. It also shows the margin between the channel level at that point and the overload point of the system. According to Holbrook and Dixon,[13] the minimum requirement in this respect is 18 db.

These results justify the assumption made in the design of the link, that intermodulation noise is negligible.

The crosstalk requirements are met in all respects.

Conclusions

The submarine-cable link between Clarenville, Newfoundland, and Sydney Mines, Nova Scotia, was completed in May 1956 and provides five carrier telephone groups, each capable of carrying twelve high-grade telephone circuits or their equivalent. The transmission objectives have been met in every respect.

Three 12-circuit groups are connected to the three groups across the Atlantic, between Scotland and Newfoundland. The other two groups are available to provide 24 circuits between Newfoundland and the mainland of Canada.

Table II

Group	Channel	Noise Level, Dba	
		Sydney Mines	Clarenville
1	1	25.0	24.5
1	12	24.5	23.2
2	1	24.0	22.5
2	12	23.5	22.0
3	1	24.0	20.5
3	12	25.0	17.5
4	1	24.5	17.5
4	12	24.5	16.5
5	1	24.5	17.5
5	12	27.0	18.0

Table III

Frequency, Kc	Equivalent at Amplifier in First Repeater		
	Channel Level, Db	Overload, Db	Margin, Db
552	−2	+20	22
312	−4	+24	28
260	−5	+25	30
20	−5	+25	30

Acknowledgments

It has been the authors' privilege to present an integrated account of the work of many of their colleagues in the Post Office and in industry. The Post Office staff have been responsible for designs and inspection, and for testing at home and in the field, as well as the laying of the submarine-cable system by *HMTS Monarch*. In Great Britain, Submarine Cables, Ltd., and the Southern United Telephone Company, Ltd., provided the submarine and overland cables respectively, while Standard Telephones and Cables, Ltd., supplied and contributed much to the design of the submerged repeaters and terminal equipment. On site, the assistance rendered by the Ordnance Survey of Great Britain; the Canadian Comstock Company, Ltd., who laid the cable across Newfoundland; the Northern Electric Company, Ltd., who carried out the terminal equipment installations; and by the other partners in the project, the American Telephone and Telegraph Company, Inc.; the Canadian Overseas Telecommunication Corporation, and the Eastern Telephone and Telegraph Company has been invaluable. The permission of the Engineer-in-Chief of the Post Office to make use of the information contained in the paper is gratefully acknowledged.

References

1. TRANSATLANTIC CABLE CONSTRUCTION AND MAINTENANCE CONTRACT, Nov. 27, 1953.

2. TRANSATLANTIC TELEPHONE CABLE SYSTEM—PLANNING AND OVER-ALL PERFORMANCE, E. T. Mottram, R. J. Halsey, J. W. Emling, R. G. Griffith. Page 842 of this issue.

3. A Transatlantic Telephone Cable, Mervin J. Kelly, Sir Gordon Radley, G. W. Gilman, R. J. Halsey. *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 124–39. (Also in *Proceedings*, Institution of Electrical Engineers, London, England, vol. 102, pt. B, Sept. 1954, p. 117.)

4. Transatlantic Telephone Cable—The Overland Cable in Newfoundland, H. E. Robinson, B. Ash. *Post Office Electrical Engineers, Journal*, London, England, vol. 49, 1956, pp. 1, 110.

5. Route Selection and Cable Laying for the Transatlantic Cable System, J. S. Jack, W. H. Leech, H. A. Lewis. Page 944 of this issue.

6. A New Key West-Havana Carrier Telephone Cable, H. A. Affel, W. S. Gorton, R. W. Chestnut. *Bell System Technical Journal*, New York, N. Y., vol. 11, 1932, p. 197.

7. Submerged Telephone Repeaters for Shallow Water, R. J. Halsey, F. C. Wright. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 101, pt. I, Feb. 1954, p. 167.

8. Modern Submarine Cable Telephony and Use of Submerged Repeaters, R. J. Halsey. *Journal*, Institution of Electrical Engineers, vol. 91, pt. III, 1944, p. 218.

9. Cable Design and Manufacture for the Transatlantic Submarine Cable System, A. W. Lebert, H. B. Fischer, M. C. Biskeborn. Page 907 of this issue.

10. Repeater Design for the Newfoundland–Nova Scotia Link, R. A. Brockbank, D. C. Walker, V. G. Welsby. Page 927 of this issue.

11. Some Experiments on the Breakdown of Heater-Cathode Insulation in Oxide-Coated Receiving Valves, G. H. Metson, E. F. Rickard, F. M. Hewlett. *Proceedings*, Institution of Electrical Engineers, vol. 102, pt. B, Sept 1955, p. 678.

12. Power-Feed System for the Newfoundland–Nova Scotia Link, J. F. P. Thomas, R. Kelly. Page 938 of this issue.

13. Load Rating Theory for Multi-Channel Amplifiers, B. D. Holbrook, J. T. Dixon. *Bell System Technical Journal*, New York, N. Y., vol. 18 1939, p. 624.

Repeater Design for the Newfoundland–Nova Scotia Link

R. A. BROCKBANK D. C. WALKER V. G. WELSBY

Synopsis: The Newfoundland-Nova Scotia cable required the provision of 16 submerged repeaters each transmitting 60 circuits in the bands 20 to 260 kc from Newfoundland to Nova Scotia, and 312 to 552 kc in the opposite direction. This paper deals with the design and production of these repeaters. Each repeater has a gain of 60 db (decibels) at 552 kc, and the amplifier consists of two forward amplifying paths with a common feedback network. Reliability is of paramount importance, and production was carried out in an air-conditioned building with meticulous attention to cleanliness and to very rigid manufacturing and testing specifications. The electrical unit is contained in a rigid pressure housing 9 feet long and 10 inches in diameter with the sea cables connected to an armor clamp and a cable gland at each end. A submerged equalizer was provided near the middle of the sea crossing.

THE BRITISH POST OFFICE has engineered many shallow-water submerged-repeater systems,[1] and there has been a progressive improvement in design techniques and in the reliability of components, which has reflected a growing confidence in the ability to provide long-distance systems having an economic life. The 7-repeater scheme from Scotland to Norway laid in 1954 introduced for the first time repeaters which would withstand the deepest ocean pressure, together with an electrical circuit which embodied improved safety and fault-localizing devices. Also, since a repeater is only as reliable as its weakest component, much greater attention and control were directed at this stage to the design, manufacture, and inspection of all components, both electrical and mechanical. This repeater design was, in fact, envisaged as a prototype for a future transatlantic project.

In the finalized plans[2] for the Newfoundland-Nova Scotia cable it was required to carry 60 circuits, so that some redesign of the 36-circuit "prototype" repeater became essential. It was accepted, however, as a guiding principle throughout the redesign that there should be no departure from previous practice without serious consideration and adequate justification. It was obviously not an occasion to experiment with new ideas.

Experiences of the Post Office and Bell Telephone Laboratories were pooled for the project, and the technical resources of both organizations were freely available at all times for consultative purposes. Detailed manufacturing and testing specifications were exchanged and approved, and each party was free to inspect the other's production methods. This mutual interchange was undoubtedly highly beneficial, and in the British case it resulted in a still more rigorous control of manufacturing and inspection methods.

Planning

General

Preliminary design calculations indicated that it should be possible to increase the circuit-carrying capacity of the Anglo-Norwegian prototype repeater from 36 to 60 circuits, and tests on a model confirmed that, with band frequencies of 20 to 264 and 312 to 552 kc, a 60-db gain at 552 kc could be realized with satisfactory margins against noise, distortion, and overload. This gain fixed the repeater spacing at about 20 nautical miles, so that on the selected route two repeaters would be required on the land section between Clarenville and Terrenceville, and 14 in the Terrenceville-Sydney Mines sea section. It was noted that the land repeaters might have to work with an ambient temperature of 12 C (degrees centigrade) higher than in the sea repeaters.

Each repeater would need to be energized with a direct current of 316 ma (milliamperes) at 124 volts, so that the total route voltage would be about 2,300 volts. This voltage would be quite acceptable to the repeaters, but for normal operation it was proposed[2] to feed from both terminals simultaneously, thereby halving the maximum voltage to ground.[3]

The precise localization of any faulty or aging repeater would be of paramount importance. It was decided to retain the two supervisory methods which on the prototype had worked satisfactorily in this respect. These consisted of a pulse-distortion equipment requiring no additional components in the repeater, and a loop-gain monitoring set involving a special unit in the repeater and the allocation of a 4-kc band (260 to 264 kc) for its operation.

Manufacture and testing of the electrical units was again to be carried out by a contractor in a temperature- and humidity-controlled production building. To enable manufacture to start as early as possible, arrangements were made for the contractor to co-operate with the Post Office at an early design stage, so that engineering could follow fast on the heels of the design. The outer housing and method of brazing-in the bulkheads had all proved entirely satisfactory on the prototype, and therefore these operations could proceed as in previous production.

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957.

R. A. Brockbank, D. C. Walker, and V. G. Welsby are with the British Post Office, London, England.

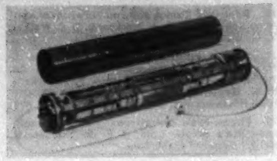


Fig. 1. Internal unit

The Post Office assumed responsibility for the production and testing of the glands, since no contractor had experience of this work.

Forward planning in early 1954 scheduled the first electrical unit to be completed in June 1955, with units following at 5-day intervals. This target was, in fact, delayed until August 1955, but all 16 working repeaters were fully tested and available before the commencement of the laying operation in May 1956.

Distortion Monitoring Equipment

The pulsed-carrier supervisory method, as used on previous systems,[1] is employed primarily for measuring the distortion on repeaters. Under normal operating conditions the distortion level may be only just noticeable above the noise but, should appreciable distortion occur, e.g., failure of one amplifying path in a repeater, it could be readily located, since the pulse amplitude from the faulty repeater would increase by about 12 db for second-harmonic distortion and about 18 db for third-order distortion.

Loop-Gain Monitoring Equipment

For the loop-gain monitoring equipment[1] the repeaters have to be designed to incorporate a second-harmonic generator operating at a frequency unique to each repeater at 120-cycle spacing in the 260- to 264-kc band. The second harmonics return to Sydney Mines in the 520- to 528-kc band, and these two channels must be taken out of service during the measurement.

Levels and Equalization

Controlling Factors

In practice, deviations from an ideal system, wherein all repeaters match the cable and operate at all times at the same levels, require the repeaters to be designed with specific margins against overload and intermodulation to meet an agreed maximum noise figure for the system under all working conditions.[2] Factors involved in assessing these margins and in planning the equalization and level diagram for the system are:

Temperature. The final, assumed sea-bottom temperature was 2.3 C, with a maximum annual variation of ±3 C. The maximum change in attenuation might therefore be ±4 db at 552 kc. The land section change would be ±3 db at 552 kc due to a possible ±10 C change on a mean of 7.5 C. The effect of these seasonal changes would be reduced by the provision of manually adjusted equalization at Clarenville, Terrenceville and Sydney Mines.

The repeaters show a small change in gain (less than 0.05 db) during the warming-up period after energization, but the effect of ambient-temperature change is negligible.

Repeater Spacing. The repeater-section cable lengths were to be cut in the cable factory in such a way that the expected attenuation at 552 kc, when laid at the presumed mean annual temperature of the location, should be 60.0 db. An anticipated decrease in attenuation of 1.42% (per cent) at 552 kc was assumed when laid. The assumed mean annual temperature of sections of the route varied between 1.7 and 4 C. Temperature corrections employed an attenuation coefficient at 552 kc of +0.16% per C. It was expected that the total error at 552 kc after laying seven repeaters would not exceed 1.5 db, and this could be largely corrected as explained in the following.

Cable Characteristics. The cable equalization built into the repeater was based on a cable attenuation characteristic which was later discovered to be appreciably different from the laid characteristic. Cutting the cable as described in the foregoing overcomes this difficulty at 552 kc, where the signal-to-noise ratio is at a minimum. The new shape of the characteristic, however, indicated that at about 100 kc the error would reach 7 db on the complete route. To reduce this deviation it was decided to introduce a submerged equalizer in the middle of the sea section to correct for half this error, and to insert, in each of the 4-wire paths of the transmit and receive equipments, equalization for one-quarter of this error. There is an appreciable signal-to-noise margin in hand at this frequency, so that the system would not be degraded below noise specification by these equalizer networks.

It was also decided that the splice at the equalizer which would connect the halves of the link together should not be completed until after the laying operation had commenced. An excess length of cable was provided on the equalizer tail, and this could be cut at a position indicated by measurements taken during the laying of the first half-section, so that the equalization at the 552-kc point could be largely corrected for laying and temperature-coefficient errors. In practice, it is not easy to separate these two factors.

Repeater Characteristic. The repeater was designed to equalize the original cable-attenuation characteristic to ±0.2 db, as this was possible with a reasonable number of components. This variation appeared as a roll in the gain/frequency characteristic, which was expected to be systematic and would therefore lead to a ±3-db roll in the over-all response. It was proposed that equalization for this should be provided at the receive terminal. Manufacturing tolerances were expected to be small and random.

Repeater Interaction. At the lower frequencies, where the loss of a repeater section is comparatively small, a roll in the over-all frequency response will arise, due to changes in the interaction loss between repeaters. The design aimed at providing a loop loss greater than 50 db, which would reduce rolls to less than ±0.03 db per repeater section, and therefore to about 0.5 db at 20 kc with systematic addition on the whole route.

Planning of Levels

From a critical examination of all these variables it was concluded that the repeater should be designed to have an overload margin of 4 db above the nominal mean annual temperature condition. It was also desirable that the system be able to operate within its noise allowance, if one path of a twin amplifier failed. Tests on a model amplifier gave overload values of +24 dbm (decibels referred to 1 milliwatt) and +19 dbm for 2- and 1-path operation respectively, so that with a single-tone overload requirement of 18 dbm[4] at a zero-level point, the maximum channel level at the amplifier output would be −3 dbr (decibels referred to reference noise) for a single amplifying path.

Thermal-noise considerations (i.e. resistance plus tube noise) fixed the minimum channel level at the repeater input at −69 dbr for the purpose of meeting the allowable system noise limit of +28 dba at a zero-level point. At 552 kc the amplifier gain is 65 db, so that the minimum level at the amplifier output is −4 dbr. A system slope of ±4 db due to temperature variations, corrected by similar networks at the transmit and receive terminal, would,

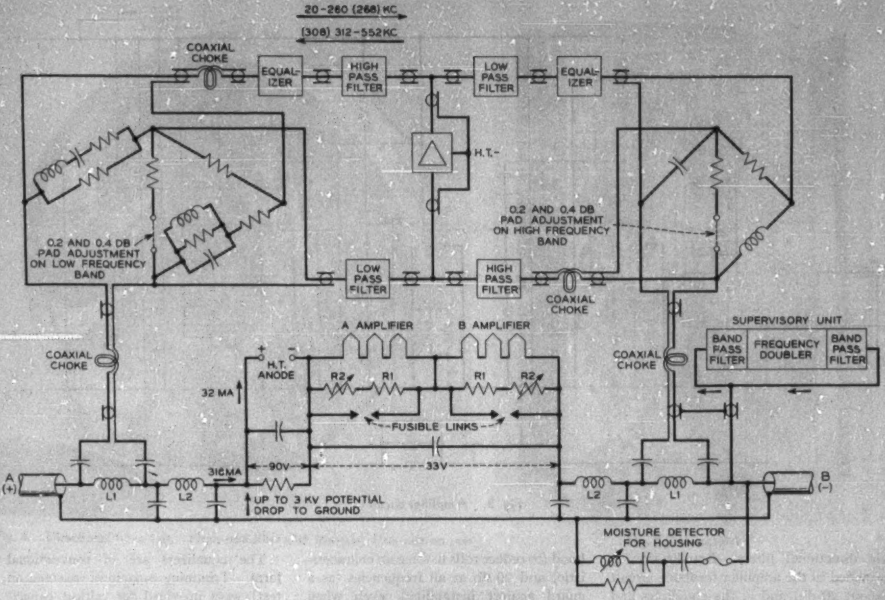

20-260 (268) KC
(308) 312-552KC
COAXIAL CHOKE
EQUALIZER
HIGH PASS FILTER
LOW PASS FILTER
EQUALIZER
H.T.-
0.2 AND 0.4 DB PAD ADJUSTMENT ON HIGH FREQUENCY BAND
0.2 AND 0.4 DB PAD ADJUSTMENT ON LOW FREQUENCY BAND
LOW PASS FILTER
HIGH PASS FILTER
COAXIAL CHOKE
SUPERVISORY UNIT
BAND PASS FILTER
FREQUENCY DOUBLER
BAND PASS FILTER
A AMPLIFIER
B AMPLIFIER
COAXIAL CHOKE
+
-
H.T. ANODE
32 MA
R2
R1
R1
R2
FUSIBLE LINKS
COAXIAL CHOKE
90V
33V
316MA
A (+)
L1
L2
UP TO 3 KV POTENTIAL DROP TO GROUND
L2
L1
B (-)
MOISTURE DETECTOR FOR HOUSING

Fig. 2. Schematic of submerged repeater

however, degrade the noise by 0.5 db. Intermodulation noise was estimated[8] on an average busy-hour basis, and it was concluded that the increase in noise at 552 kc from this source was negligible: less than 1 db, even with several repeaters in which the amplifier had failed on one path. At lower frequencies the contribution from intermodulation noise is greater, and at 20 kc it exceeds resistance noise. However, at 20 kc the total noise is some 8 db below the specification limit, and therefore again several amplifiers could fail on one path before the noise exceeded the specification limit. Actually it was discovered that the predominant source of third-order intermodulation on the repeater was in the nickel-iron/ceramic seals on high-voltage capacitors and followed a square law with input levels.

From a more detailed examination of the factors briefly mentioned in the foregoing, it was decided that the initial line-up should be based on a nominal flat −3.5-dbr point at the amplifier output and the final working levels decided upon as the results of tests on the completed link.

With equal loading on the grid of the output tube at all frequencies, the worst signal-to-noise ratio exists at 552 kc; some pre-emphasis on the transmit signal should therefore prove beneficial. In fact, after completing the tests on the link, it was decided to improve the margin on noise by raising the level at 552 kc by 2 db, thus giving a sloping level response at the amplifier output in the high-frequency band. To maintain the same total power loading, the low-frequency band levels were decreased by one db, still retaining a flat response.

LAYING

It was proposed to use laying methods with continuous testing similar to those employed successfully on the Anglo-Norwegian project. The complete link with a temporary splice at the equalizer would be assembled and tested on board *HMTS Monarch* and laying would proceed from Terrenceville to Sydney Mines in the high-frequency direction of transmission. A detailed description of the actual laying operation is given elsewhere.[9] After completion of tests on the submarine section, the land section to Clarenville would be connected with appropriate equalization at Terrenceville.

Design of Electrical Unit of Submerged Repeater

GENERAL

The electric equipment is contained in a hermetically sealed, brass cylinder (filled with dry nitrogen) $7^3/_4$ inches in diameter and 50 inches long, which is bolted at one end to one of the bulkheads of the housing. A flexible coaxial cable emerges through an O-ring seal at each end, and these are ultimately jointed to the cable glands. The various units forming the complete electrical unit are mounted within a framework of Perspex (polymethylmethacrylate) bars which forms the main insulation of the repeater, and these units may operate at 3-kv d-c to the grounded brass cylinder. Fig. 1 shows the construction.

A schematic of the electric circuit is given in Fig. 2. The direct current for energizing the repeater is separated from the carrier transmission signals by the *A*- and *B*-end power-separating filters, and passes through the amplifier tube heaters and a chain of resistors which develops the 90-volt high-voltage supply for the amplifier. The carrier-frequency signals pass through the same amplifier

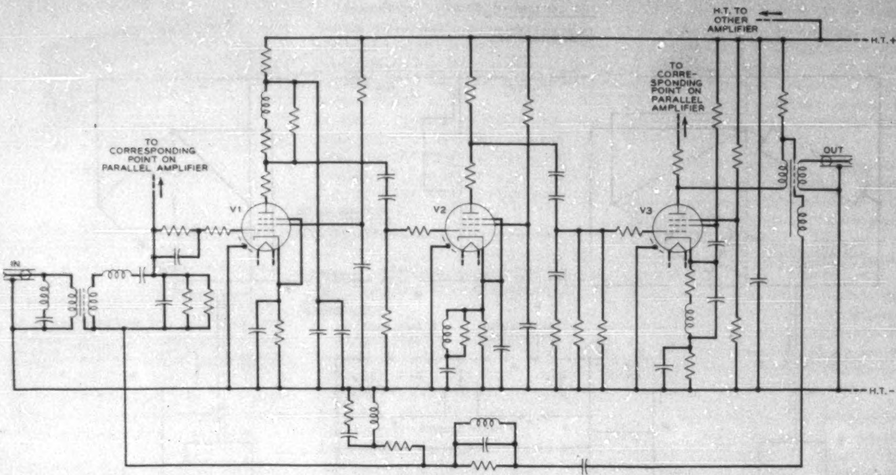

H.T. TO OTHER AMPLIFIER
H.T.+
TO CORRE-SPONDING POINT ON PARALLEL AMPLIFIER
OUT
TO CORRESPONDING POINT ON PARALLEL AMPLIFIER
V1
V2
V3
IN
H.T.-

Fig. 3. Amplifier circuit

via directional filters. Equalization is provided in the amplifier feedback circuit (about 20 db) and in the equalizers and the bridge networks which combine the directional filters. The main purpose of the bridges, however, is to reduce the severe harmonic requirement on the directional filters due to having high- and low-level signals present at the repeater terminals. The whole carrier circuit is designed on a nominal impedance of 55 ohms. Attached to the *B*-end of the repeater is the loop-gain supervisory unit and also, via a high-voltage fuse, a moisture-detector unit which is used primarily during the high-pressure test to confirm that the housing is free from leaks. The latter comprises a series-resonant circuit at about 1.3 mc (megacycles), in which the inductance is varied by the gas pressure on an aneroid capsule mounted in the space between the electrical unit and the housing. The presence of moisture in this cavity increases the gas pressure owing to the release of hydrogen by the reaction of water vapor with metallic calcium held in a special container. At a later stage, the fuse is blown to disconnect this circuit.

The circuit design of the repeater introduces multiple-shunt paths across the amplifier, and care has to be taken to ensure that there is adequate attenuation in each path. In general, the design is such that the combination will give a loop loss of a least 40 db in the working band (to reduce rolls in the gain characteristic) and 20 db at all frequencies (as a guard against instability), even when one repeater terminal is open- or short-circuited to simulate a faulty cable.

UNIT DETAILS

Power Filter

The power filters are in effect a series pair of high- and low-pass filters (see Fig. 2). The shunt capacitors may have to withstand 3 kv, and clearances on the input cable and some wiring have to be adequate for this voltage. The inductors have to carry the line current of 316-ma d-c, and the intermodulation must be extremely low (see section on inductors).

Directional Filter

The directional filters are a conventional Zobel high-pass and low-pass filter pair with a susceptance-annulling network. Silvered-mica capacitors and carbonyl-iron dust-cored inductors are used. The bridges combining the "go" and "return" filters reduce the distortion, caused by the ferromagnetic material, to an acceptable level.

Bridge and Equalizer

A simple nonresonant bridge is used at the *B*-end of the repeater, but the *A*-end bridge is a resonant type and provides a substantial degree of equalization (see Fig. 2).

The equalizers are of conventional form. Trimming capacitors (selected on test) were provided for critical capacitances in order to utilize standard tolerances on all capacitors. A pad of 0.2 db and 0.4 db is provided on each equalizer unit, so that the repeater low-frequency or high-frequency path can be independently trimmed to give the best match to the target response for the repeater.

The components in this circuit were small air-cored inductors, silvered-mica capacitors, and wire-wound resistors, except for a few high-resistance ones, which were of the carbon-rod type. Included in this unit are coaxial chokes which separate parts of the circuit to avoid the effect of multiple grounding. They are merely inductors wound with coaxial wire on 2-mil permalloy-C tape ring cores.

Supervisory Unit

The supervisory unit comprises a frequency-selection crystal filter of about 100-cycle bandwidth in the range of 260 to 264 kc, fed from the low-frequency output end of the repeater via a series resistor. This filter feeds a full-wave germanium point-contact crystal-rectifier bridge which acts as a frequency doubler. The second harmonic in the 520- to 528-kc band is filtered out by a coil-capacitory band-pass filter, and fed back through a resistor to the same point

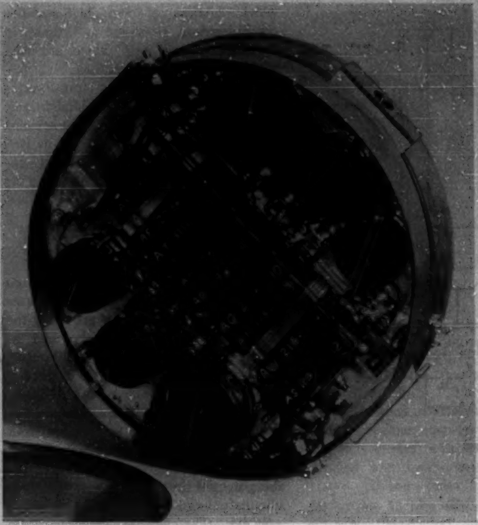


Fig. 4. Directional filter unit. High-pass filter with low-pass filter can on rear

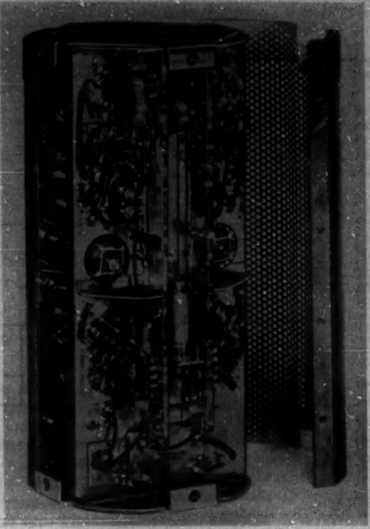


Fig. 5. Amplifier

in the repeater. The two series resistors minimize the bridging loss of the unit on the repeater and ensure that a faulty supervisory component has negligible effect on the normal working of the repeater.

D-C Path

The d-c path includes a resistor providing the 90-volt supply and the heater chain of six electron tubes (see Fig. 2). The voltage drop across the heater chain is not utilized for the amplifier high-voltage supply, as the heaters would then be at a positive potential with respect to the cathodes, thereby increasing the risk of breakdown of heater-cathode insulation. There would also be a complication in maintaining the constant heater current, particularly if the high-voltage supply current fail in one path of the amplifier. The normal amplifier high-voltage supply current is 32 ma.

It is essential to maintain a d-c path through the repeater even under fault conditions, in order that fault-location methods may be applied. Special care has therefore been taken to provide parallel paths capable of withstanding the full line current. For example, the high-voltage resistor actually consists of a parallel-series combination of ten resistors, and the whole assembly is supported on Sintox (sintered alumina) blocks which maintain a good insulation at 3-kv d-c, even at high temperatures.

Electron tube operation for consistent long life indicates the necessity of maintaining a specific constant cathode temperature; to achieve this, electron tubes are arranged according to heater characteristics into six heater-current groups between 259 and 274 ma and stabilized to ±1%. The appropriate heater-shunt resistor is applied so that the tube operates correctly with 316-ma line current, but for convenience the shunt is taken across each set of three tubes, all in one heater group, forming one amplifier path. *R*1 is fixed (300 ohms) and *R*2 is selected to suit the tubes. *R*1 is the resistance winding of a special short-circuiting fuse. When energized by the full line current should a heater become open-circuited, it causes a permanent direct short-circuit across the heater chain. The line voltage will be temporarily increased by about 95 volts while the fuse operates (1 minute) and will then drop to 12 volts below normal.

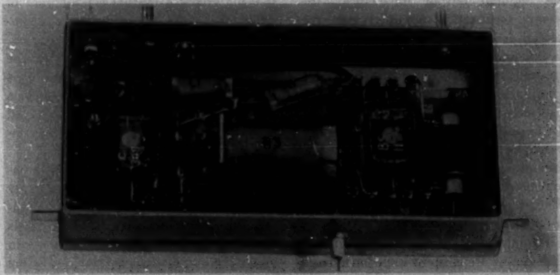


Fig. 6. Feedback unit of amplifier

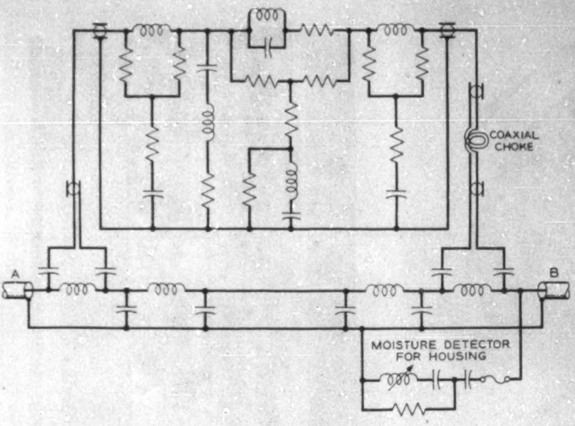


Fig. 7. Schematic of submerged equalizer

Amplifier

The amplifier circuit is shown in Fig. 3. It consists of two 3-stage amplifiers connected in parallel between common input and output transformers with a single feedback network. This circuit arrangement allows one amplifier path to fail without appreciably affecting the gain of the complete amplifier (less than 0.1 db for all faults except those on the grid of *V*1 and the anode of *V*3), but the overload point is reduced by about 5 db and distortion at a given output level is increased (about 12 db for second harmonics). Care has been taken to ensure that the open- or short-circuiting of a component in one amplifier path will not affect the performance, life, or stability of the remaining path, and this involves the duplication of certain components.

Mixed feedback is employed to produce the required output impedance; the current feedback is obtained from the resistor feeding the high-voltage supply to the output transformer, and the voltage feedback is developed across a 2-turn winding on the output transformer, which also serves as a screen. The output of the feedback network is fed in series with the input signal to the grid of *V*1. The gain response of the amplifier is chiefly controlled by the series-arm components in the feedback network, which resonate at 600 kc.

The input transformer is built out as a filter and it steps up in impedance from 55 ohms to 17,000 ohms. Protective impedances minimize the effect of a short-circuit on the grid of one of the first-stage tubes. The anode load of the first stage resonates at 600 kc, and is roughly the inverse of the feedback network, giving constant feedback loop gain over the working frequency band. The output tube has about 5.5 db of feedback from its cathode resistor, and the pair of output tubes feeds the output transformer, which steps down from 5,000 ohms to 55 ohms.

Specially designed, long-life tubes are used.[7] The first two stages are operated at about 40 volts on the screens and anodes; each anode current is 3 ma, giving a mutual conductance of 5.1 ma per volt. The output stage is operated at 60 volts on the screen, at +15 volts on the suppressor grid to sharpen the knee of the V_a/I_a characteristic, and at nearly the full high-voltage supply of 90 volts on the anode; the anode current is 6 ma giving a mutual conductance of 6.6 ma per volt. The tube dynamic impedance is approximately 300,000 ohms. To obtain an anode current nearest to the design value (and one for which the tubes are aged), one of two values of bias resistance can be selected for *V*1 and *V*2, and one of three values of bias resistance for *V*3.

All capacitors subject to the high-voltage supply are of the oil-filled paper type, and the others are of the silvered-mica type. Inductors are air-cored spools which are multisectioned when used in high-impedance circuits. All resistors are of the solid carbon-rod type, except for the input-transformer termination, which consists of two high-stability cracked-carbon resistors, and those in the feedback network, which are wire wound.

The input and output transformers employ 2-mil permalloy-C laminations, and the latter core is gapped on account of the polarizing current. A narrow Perspex spool fits the center limb, and conventional layer windings are used; the screen is a sandwich made of copper foil and adhesive polythene tape.

MECHANICAL DESIGN DETAILS

The general arrangement may be seen in Fig. 1. At the *A*-end is a cast-brass pot containing the resistors providing the amplifier high-voltage supply, and as this is bolted directly on to the housing bulkhead, the heat generated is readily conducted away. The remaining units are in cylindrical cans mounted in the insulating framework formed by four Perspex bars. These are sprayed with copper on both faces to guarantee the d-c potential on these surfaces and they eliminate the risk of ionization at working voltages. The cans are not hermetically sealed but are dried out with the repeater when it is finally sealed and filled with dry nitrogen. Perforated covers on the amplifier allow air circulation to reduce the ambient temperature.

Fig. 4 shows a typical can assembly, and Figs. 5 and 6 the construction of the amplifier. It will be seen that the latter is a double-shelved structure with tubes alternating in direction, and an amplifier path located on each side of the chassis; the input and output transformers are at opposite ends, and the feedback network is contained in a hermetically sealed can in the center of the unit.

All cans are finished with a gold flash which is inert and, giving a clean appearance, stimulates a high standard of workmanship. Tin plating was formerly used, but it has been shown that tin tends to grow metallic whiskers.[8] Unfortunately, certain capacitor cans had to be tin plated, which necessitated extra precautions in wide clearances or protective shields. The risks from growth on soldered surfaces are not thought to be great, as all solders used have less tin content than the electric alloy. Gold plating is used for all connecting wires instead of the usual tin plating.

The insulating subpanels in units are usually made of Perspex, but where the items are subjected to high temperatures (e.g., the resistance box), Sintox, a sintered alumina ceramic, or Micalex, is used.

Polytetrafluoroethylene (ptfe) is another insulant used, and ptfe-covered

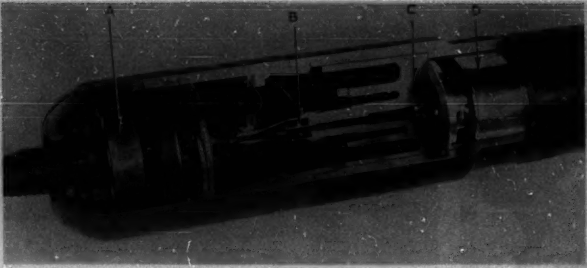


Fig. 8. Details of housing construction

A—Armor clamp B—Gland C—Bulkhead D—Part of internal unit

wire, threaded through copper tubing forms the coaxial interconnecting leads between the can units.

Careful attention is paid to the mounting of components. Small resistors, etc., are supported by soldering to tags which are the appropriate distances apart, and multilimb tags are employed to minimize the number of soldered joints. Where it is essential to solder more than one wire per limb on a tag, they must be soldered at the same time. Larger components are clamped. Electron tubes are mounted in a holder to facilitate preliminary testing with "standard electron tubes," and they have a sprung-nylon retainer; the final electrical connection is made by soldering on to an extension of the wire leading through the pins.

Design of Internal Unit of Submerged Equalizer

The submerged equalizer corrects for the difference between the assumed design cable characteristic and the subsequently determined laid characteristic for equal attenuation lengths at 552 kc. It also absorbs the loss of 9 nautical miles of cable and has an attenuation of 26.0 db at 552 kc.

The construction is identical with the submerged repeater except for the replacement of all can assemblies, other than the power filters, by the equalizer cans. Fig. 7 is a schematic of the unit.

Electrical Components

GENERAL

The components used in the repeaters were either designed specially for submerged repeaters or were standard items with improvements. There are approximately 300 components in each repeater, of which 110 are in the amplifier. Rigorous control of manufacture and meticulous inspection are imperative to ensure a consistent long-life product, and cleanliness is essential at all stages. In some cases "belt-and-braces" techniques can be effectively employed, e.g., by using double connections. Over 1,500 separate soft-soldered connections are involved in the complete assembly. Much work has been done on components, but only a brief indication can be included here. A range of typical components appears in Figs. 4–6.

RESISTORS

Resistors fall into the following categories:

1. Power resistors used solely for d-c purposes (e.g., resistors providing the amplifier high-voltage supply). These are wire-wound vitreous-enamelled resistors on Sintox ceramic formers. Nichrome terminal leads are used, and all connections are brazed.

2. High-frequency resistors, whose tolerance is not close, and which often carry direct current but of low power (e.g., anode load resistances). A modification of a standard carbon-rod resistor is used. The ends of the rod are copper plated and the end caps and terminal leads are soldered on. The tolerance is normally ±5%, and the maximum rating permitted is about one quarter of the commercial rating.

3. Precise high-frequency resistors of resistance below 1,000 ohms (e.g., feedback components). Here wire-wound spool resistors are suitable, and bifilar or reverse layer windings with Lewmex enamel and silk-covered wire are used.

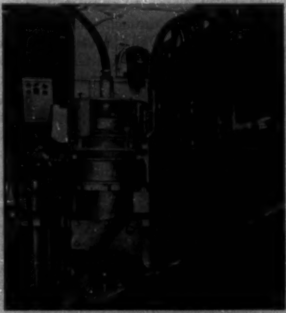
4. Precise high-frequency resistors of high resistance. For terminating the input transformer a resistance of 17,000 ohms is required. Because it is not possible to make a suitable wire-wound resistor, high-stability cracked-carbon film resistors are used, but to minimize the effect of a disconnection two are used in parallel.

INDUCTORS

The majority of inductors used in the amplifier and equalizer do not require a high Q-factor. They are wound on air-cored ceramic bobbins of four types, and the high-inductance ones are sectionalized. In general, solid wire with Lewmex enamel and double-silk covering is used for the amplifier inductors, and stranded wire for the equalizer inductors.

A high Q-factor inductor is essential in the directional filters, and a carbonyl-iron pot core was used; the Q-factor is about 250 at 300 kc. Precise adjustment and stability of inductance were obtained by setting the gap between the halves of the pot core with a cement of Araldite (an epoxy resin) and titanium dioxide. A Perspex former was used.

Special inductors were required in the power filters to take the 316-ma d-c line current. A wave-wound air-cored coil was used for the carrier-path filter (L_1 in Fig. 2), and a toroid on an audio-frequency Permalloy dust-core ring was used for the low-pass filter (L_2 in Fig. 2). Solid wire, with Lewmex and double-silk covering, was used to keep the d-c resistance to a minimum.



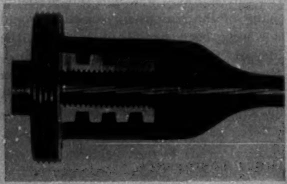


Fig. 9 (left). Silver-soldering of bulkhead into housing with induction heater

Fig. 10 (above). Section of high-pressure gland

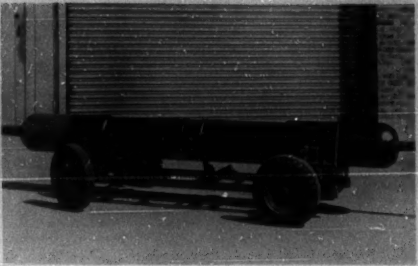


Fig. 11 (left). Coil production in the "dairy"

Fig. 12 (above). Completed repeater

CAPACITORS

Capacitors are divided into three categories:

1. Those subjected to the full line voltage which may operate at up to 3 kv.
2. Those subjected to the amplifier high-voltage supply of 90 volts.
3. Those which have negligible polarization (less than 10 volts), and which are often required to precise values.

Categories 1 and 2 are of the oil-filled paper type, with respectively four layers of 36-micron and three layers of 7-micron kraft paper. The oil is a mineral type loaded with 18% resin, and the capacitors are filled at 60 C and sealed at room temperature.

Small capacitors and those of precise value as in category 3 are silvered-mica capacitors. These are encased in an epoxy resin to give mechanical protection and a seal against moisture. Visual inspection of all mica plates is made before and after silvering, and any with cracks, inclusions, stains, or any other abnormality are rejected. Mica is a very variable material and at times the percentage of rejects was high, but probably many of the reasons for rejection would not have been significant as far as the life of the capacitor was concerned. However, experience has shown that even with stringent precautions mica is not an entirely satisfactory dielectric material.

OTHER COMPONENTS

Electron tubes form the subject of a separate paper.[7] Of the other miscellaneous components used, one of interest is the short-circuiting fuse across the electron tube heaters. Its construction is like that of a normal wire-wound resistor on a Sintox tube former, but inside are two cupped copper electrodes filled with a low-melting-point eutectic alloy. If the full line current (316 ma) is passed through the winding, owing to a heater disconnection, the heat generated is sufficient to melt the alloy, which then fuses the two electrodes together. The winding is thus short-circuited, and a permanent connection is left between the electrodes.

Design of Housing and Gland

GENERAL

Although the maximum depth of the water in which British rigid-type repeaters were laid did not exceed about 250 fathoms, the housings used for these repeaters were generally of a type designed for use at ocean depths, and when connected into the cable they were amply strong enough to transmit stress up to the breaking point of any of the cables used.

The part of the housing which is sealed against water pressure consists essentially of a hollow cylinder, machined from hot-drawn steel tube, and closed at both ends by steel bulkheads carrying the cable glands, through which the connections are made to the electrical unit (see Fig. 8). The latter is bolted rigidly to the inner face of the *A*-end bulkhead. The steel blanks used for the main cylinder and the bulkheads are tested with an ultrasonic crack detector, and after machining they are further subjected to magnetic crack-detection tests.

Each gland has a brass cover which completes the coaxial transmission path and contains a weak solid mixture of polythene and polyisobutylene (pib). Outside the brass cover is a larger chamber closed by a flexible polyvinylchloride (pvc) diaphragm and containing pib, a viscous liquid, which prevents sea water from coming into direct contact with the bulkhead seal and the gland assembly.

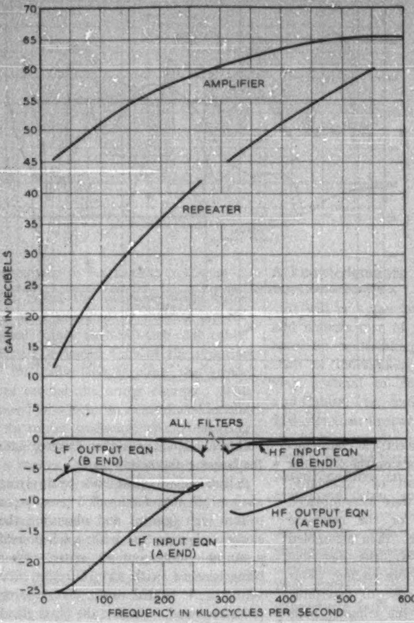
Cylindrical extension pieces, screwed on to the main casing, contain the clamps for attaching the repeater housing to the armor wires of the sea cable, and the housing is completed by dome-shaped end covers. Two external annular ridges near the center of the housing accommodate the special quick-release clamp used for handling the repeater during the laying operation.

Protection against corrosion is provided by shot-blasting the surface and then applying hot-sprayed zinc to a thickness of 0.010 inch, followed by two coats of vinyl paint. The *A*-end of the repeater is finished in red. The dimensions of the complete repeater are 8 feet $11^{3}/_{8}$ inches by $10^{1}/_{2}$ inches in diameter, and its weight is 1,150 pounds in air.

SEALING OF HOUSING

The bulkheads, which register on seatings designed to withstand the axial thrust from the water pressure, are in the form of disks with extended skirts. A watertight and diffusion-proof seal is formed between the casing and the outer skirt of each bulkhead by a silver-soldering process, with the use of carefully controlled electromagnetic induction heating to raise the jointing region to the required temperature. The diametral clearance between the cylinder and the locating surface of the bulkhead is 0.003 ±0.002 inch, the diameter of the bulkhead being reduced by 0.004 inch for an axial distance of 3 inches from the rim of the skirt to provide a recess into which the molten solder can flow.

The solder is applied as eight preformed No. 16 standard wire gauge wire rings



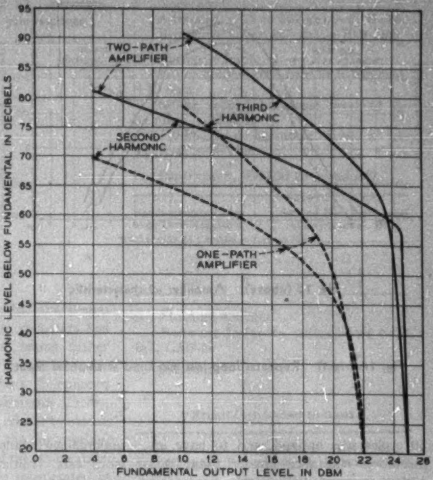


Fig. 13 (left). Repeater gains and losses

Fig. 14 (above). Amplifier distortion

which are fitted into place when cold, and coated with a paste formed by mixing flux powder with dehydrated ethyl alcohol. The generator used for heating has a nominal output of 50 kw at a frequency of about 350 kc and is capable of raising the temperature of the jointing region to 750 C in 5 minutes. The temperature, as indicated by four thermocouples inserted in special holes drilled in the ends of the casing, is maintained at 750 C for a period of 45 minutes to allow ample time for the entrapped gas and flux pockets to float to the surface. Fig. 9 shows the arrangement for soldering in a bulkhead.

The primary object of the outer skirt is to keep the heated region far enough away from the base to prevent the temperature of the latter from rising unduly. Temporary water jackets are also clamped over the gland and around the outside of the casing during the sealing operation. A subsidiary skirt on each bulkhead contains a vent hole which serves to allow displaced air to escape as the bulkheads are inserted into the casing. These vents are later used to apply a low-pressure gas-leak test to the bulkhead seals, and then to flush the housing with dry nitrogen to remove any trapped moisture. The vents are finally sealed. At this stage, the sealed housing is pressure-tested in water at $1\frac{1}{2}$ long tons per square inch for a period of 7 days, a moisture detector (mentioned previously) being used to check the leakage. (One long ton = 2,240 pounds.)

Glands

The deep-sea gland was developed from the castellated gland which has been used successfully for a number of years in shallow-water repeaters. The basic principle of this gland is very simple and is shown in Fig. 10; the polythene-insulated cable core passes right through the bulkhead, and the initial seal is formed by the contraction, during cooling, of polythene molded on to the core and enclosing a steel stem, having a castellated profile, forming part of the bulkhead. Each side of the castellation has a taper of about 7 degrees. The application of water pressure increases the contact pressure between the polythene and the stem, making the gland inherently self-sealing. As an additional safeguard, the gland stem is first prepared by a lead plating and anodizing process, followed by the application of a thin film of polythene which forms a chemical bond to the plated surface. The injected polythene merges with this film, thus bonding the molded portion to the castellated stem. Tests on sample glands, with radioactive tracer, have shown no measurable) water diffusion (less than 0.001 milligram) at a pressure of 5 long tons per square inch over a period of 6 months.

Intrusion of the polythene into the housing, at hydraulic pressures up to at least 6 long tons per square inch, is eliminated by the use of a small-diameter core and by the provision of a screw thread in the hole through which the core passes. During the molding operation a corresponding thread is formed on the polythene core itself. This method of construction distributes the axial force over a sufficiently wide area to prevent any appreciable creep of the polythene.

The completed gland assemblies were all subjected to a minute X-ray examination, followed by a pressure test at 5 long tons per square inch for 3 months. While under pressure, the glands had to withstand a voltage test of 40 kv d-c for one minute, to show no ionization effects when a voltage of 3 kv (rms) at

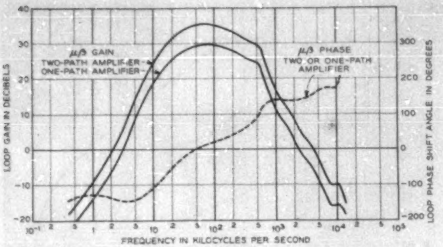


Fig. 15 (above). Amplifier $\mu\beta$ characteristic

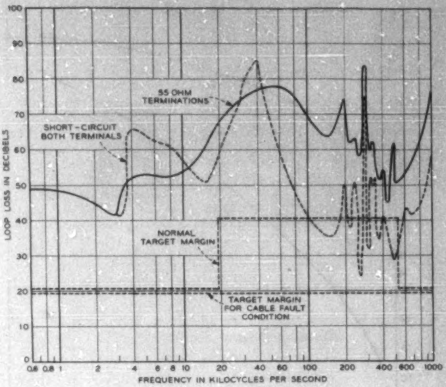


Fig. 16 (right). Repeater loop-loss measured at amplifier input

50 cycles was applied, and to have an insulation resistance greater than 20×10^{12} ohms.

Manufacture

General

The manufacture of the electrical units was carried out in accommodation specifically designed for submerged-repeater production. Temperature was controlled at 68 F (degrees Fahrenheit) and the relative humidity was less than 20% in the component shops and 40% in the assembly and test shops. Filtered air forced into the building maintained a slight positive pressure with respect to the outside and eliminated the ingress of dust. With the exception of the tubes and some resistors, all components were manufactured in this "dairy" (see Fig. 11). Operators are specially selected, and they must change into clean protective clothing in an anteroom before entering the working area, where no smoking or eating is permitted. All operators are particularly encouraged to report or reject any condition which is abnormal or in which they have not complete confidence. Rigorous inspection and testing were carried out by the contractor at all stages, and a Post Office team collaborated with "floor" inspection and the examination and approval of test results.

The Electrical Unit

The components, after the most careful examination and testing (which in some cases included an aging test), were assembled into their cans and then subjected to a shock test before undergoing detailed electrical-characteristics tests. Initial tests on the amplifier were done with a set of "standard" electron tubes, which were later replaced by the final tubes for the complete tests. The cans were then assembled in a repeater chassis; the electrical tests were required before sealing. This included a gain response to determine the best settings for the trimmer pads in each transmission band. After fitting and sealing the outer brass case, dry nitrogen was blown through the unit for 24 hours, then the gas holes were sealed. The over-all electrical characteristics were then taken, the repeater was energized for a 2-week "confidence trial" and the characteristics were rechecked. During the confidence trial, the gain was continuously monitored on a recorder (duplicated to distinguish between test equipment and repeater variations) on which changes in gain of 0.01 db were clearly indicated. On satisfactory completion of these tests the unit was ready for housing.

Assembly of Electrical Unit in Housing

The first step was to complete the molded joint between the tail cables from the *A*-end of the electrical unit and the low-pressure side of the appropriate bulkhead. This joint was X-rayed and proof-tested at 20 kv d-c for 1 minute. The electrical unit was then bolted to the bulkhead, the slack tail cable being correctly coiled into the recess provided, and the whole assembly was lowered into the housing for the first silver-soldering operation. Following this sealing, the tail cable joint was made to the *B*-end bulkhead, which was then lowered into the housing and sealed.

A leak test was then made by applying an internal air pressure of 5 pounds per square inch (gauge) and observing the surface when wetted with a solution of a suitable detergent in water. After being flushed with dry nitrogen, the vents were sealed and the housing pressure-tested. Finally, the brass gland covers were fitted and filled with compound, the extension pieces were screwed on, the flexible diaphragms were fitted, and the internal space was filled with polyisobutylene. The housing was then ready for further electrical testing.

Tests on Complete Repeater

After housing, the repeaters were submerged in a tank of water for a 3-month electrical confidence trial. Before and after the trial, the complete characteristics were checked and the noise was monitored on both terminals; during the first and last few weeks of the trial the gain was monitored on recorders in both directions. Owing to the insertion and withdrawal of repeaters in the power circuit from time to time, the repeaters were subjected to several power-switching operations and temperature cycles.

Stability of electrical characteristics, particularly gain, between the prehousing tests and the completion of the confidence trial some 4 months later was regarded as an important criterion of the reliability of a repeater. Unfortunately, test conditions, in addition to differences between and stability of the testing equipments, reduced the accuracy originally expected; nevertheless, changes of over 0.1 db were regarded as significant.

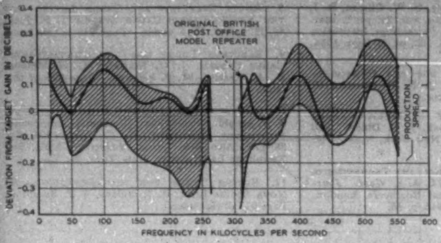


Fig. 17. Repeater-gain response

CONNECTION OF REPEATERS TO CABLE

On board the cable ship the cable ends were prepared by making tapered molded joints to 0.310-inch tail cable, sliding the domed ends of the repeater up the cable and forming the armor wires around the armor clamps. The tapered joint included a castellated ferrule on the center conductor, which, operating on the principle of the main gland, acts as a barrier against the possible passage of water down the center conductor into the repeater. The final assembly operation consists of jointing the tail cables, bolting the armor clamps to the repeater housing, and screwing on the domed ends. Fig. 12 shows a completed repeater connected to a cable.

Performance

The gains and losses of various sections of the repeater are shown in Fig. 13. Figs. 14 and 15 show respectively the harmonic distortion and the stability characteristics of the amplifier with one and two paths operating. The total shunt loss across the amplifier is shown in Fig. 16, as a margin above the amplifier gain. The curves show the result with 55-ohm terminations on the repeater and with a short circuit on each terminal. Fig. 17 shows the production spread in gain of the 16 repeaters for the system as a deviation from the target value. The highest standard deviation (at 260 kc) was only 0.11 db. In all respects the production repeaters proved to be very consistent and satisfactory in their performance, and differed little from the original laboratory-built model. Typical electrical characteristics of a repeater and the submerged equalizer are shown in Appendixes I and II respectively.

The characteristics of the completed link are described elsewhere,[9] but it is of interest that the over-all tests showed that the link behaved as predicted and met the noise requirement and the design margins.

Acknowledgments

It will be appreciated that the design and manufacture of these repeaters has been an undertaking of teams rather than of individuals. The authors are very grateful to Standard Telephones and Cables, Ltd., and Submarine Cables, Ltd., for their unsparing efforts to ensure the highest standards in the engineering, production, and testing of these repeaters. Numerous firms have also co-operated in specialized fields, and the authors are very appreciative of their help. The authors also wish to acknowledge the enthusiastic assistance of their many colleagues in the Research Branch of the Post Office in the design and in the supervisory inspection during manufacture. Finally a grateful acknowledgment is made to the engineer-in-chief of the British Post Office for permission to publish the paper.

Appendix I. Performance of Typical Submerged Repeater

1. Insulation resistance, 8,000 megohms (cold)
2. D-c resistance at 20 C

Current, Ma	D-C Resistance, Ohms
5	343.0
20	343.2
50	344.2
100	347.9

3. Voltage drop at 316 ma, 124 volts
4. Carrier gain (55 ohms) without moisture detector

Frequency, Kc	Gain, Db	Frequency, Kc	Gain, Db
20	11.54	308	44.52
30	13.82	312	44.87
50	17.47	320	45.56
100	25.06	330	46.29
150	30.86	350	47.75
200	35.81	400	51.19
230	38.40	450	54.20
260	40.96	500	57.31
264	41.13	552	60.01
268	40.83		

5. Noise level
   A-terminal (312 to 552 kc), −59.8 dbm
   B-terminal (20 to 260 kc), −70.3 dbm
6. Harmonic level
   170-kc fundamental level at B-terminal, +10 dbm
   340-kc second harmonic level at A-terminal, −60 dbm
   510-kc third harmonic level at A-terminal, −58 dbm
7. Supervisory, 260.800 kc (nominal)

Fundamental Level at B-Terminal, Dbm	Second Harmonic Level at A-Terminal, Dbm
−12	−26
−2	−13.8
+8	−3.6

8. Moisture detector
   Resonant frequency with 30-foot cable tail, 1,237 kc
9. Impedance

Return Loss Against 55 Ohms

Frequency, Kc	A-Terminal Return Loss, Db	B-Terminal Return Loss, Db
20	17	8
50	17	13
100	16	15
200	16	4
260	10	4
312	13	21
350	14	14
500	18	16
552	16	25

Appendix II. Performance of Submerged Equalizer

1. Insulation resistance, 8,000 megohms
2. D-c resistance at 20 C, 9.2 ohms
3. Voltage drop at 316 ma, 3.0 volts
4. Carrier loss (55 ohms), without moisture detector

Frequency, Kc	Loss, Db	Frequency, Kc	Loss, Db
20	4.28	260	15.90
30	4.23	312	18.26
50	4.87	350	19.96
100	7.47	400	21.75
150	10.35	450	23.16
200	13.04	500	24.55
230	14.52	552	25.97

5. Moisture detector
   Resonant frequency with 5-foot cable tail, 1,387 kc
6. Impedance
   Return loss against 55 ohms

Frequency, Kc	A-Terminal Return Loss, Db	B-Terminal Return Loss, Db
20	35	31
50	19	19
100	25	25
260	34	27
552	30	23

References

1. SUBMERGED TELEPHONE REPEATERS FOR SHALLOW WATER, R. J. Halsey, F. C. Wright. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 101, pt. I, Feb. 1954, p. 167.

2. THE NETHERLANDS-DENMARK SUBMERGED REPEATER SYSTEMS, A. H. Roche, F. O. Roe. *Ibid.*, p. 180.

3. THE BRITISH POST OFFICE STANDARD SUBMERGED REPEATER SYSTEM FOR SHALLOW WATER CABLES (WITH SPECIAL MENTION OF THE ENGLAND-NETHERLANDS SYSTEM), D. C. Walker, J. F. P. Thomas. *Ibid.*, p. 190.

4. A TRANSATLANTIC TELEPHONE CABLE, Mervin J. Kelly, Sir Gordon Radley, G. W. Gilman, R. J. Halsey. *AIEE Transactions*, vol. 74, pt. I, Mar. 1955, pp. 124–39. (Also in *Proceedings*, Institution of Electrical Engineers, vol. 102, pt. B, Sept. 1954, p. 117.)

5. POWER-FEED SYSTEM FOR THE NEWFOUNDLAND-NOVA SCOTIA LINK, J. F. P. Thomas, R. Kelly. Page 938 of this issue.

6. LOAD RATING THEORY FOR MULTI-CHANNEL AMPLIFIERS, B. D. Holbrook, J. T. Dixon. *Bell System Technical Journal*, New York, N. Y., vol. 18, 1939, p. 624.

7. NON-LINEAR DISTORTION IN TRANSMISSION SYSTEMS, R. A. Brockbank, C. A. A. Wass. *Journal*, Institution of Electrical Engineers, London, England, vol. 92, pt. III, 1945, p. 45.

8. ROUTE SELECTION AND CABLE LAYING FOR THE TRANSATLANTIC CABLE SYSTEM, J. S. Jack, W. H. Leech, H. A Lewis. Page 944 of this issue.

9. ELECTRON TUBES FOR THE TRANSATLANTIC CABLE SYSTEM, J. O. McNally, G. H. Metson, E. A. Veazie, M. F. Holmes. Page 898 of this issue.

10. METAL WHISKERS—A FACTOR IN DESIGN, S. N. Arnold. *Proceedings*, Electronic Components Symposium, Washington, D. C., May 1954, pp. 38–44. (Also *Bell System Monograph 2338*, Bell Telephone Laboratories, New York, N. Y.)

11. SYSTEM DESIGN FOR THE NEWFOUNDLAND-NOVA SCOTIA LINK, R. J. Halsey, J. F. Bampton. Page 917 of this issue.

Power-Feed System for the Newfoundland–Nova Scotia Link

J. F. P. THOMAS   R. KELLY

Synopsis: Design engineers now have available the results of many years of operating experience of submerged-repeater systems supplied from electronic, electromagnetic, and rotary-machine power equipments. To meet the unusually high standards of reliability required for the transatlantic telephone system, a scheme has been evolved that is a combination of new developments and the best features of previous methods. Electronic electromagnetic equipment forms the basis of an automatic no-break system requiring little routine maintenance.

THE OPERATING power for the submerged repeaters of the Clarenville–Sydney Mines link is derived from a constant current supplied over the central conductor from power equipments located at the two terminal stations. To protect the electron tubes in the repeaters, the current must be closely maintained at the design value, irrespective of changes in the main supply voltage or ground potential differences between the two ends of the link. Automatic tripping equipment must be provided to disconnect the cable supply should the current deviate beyond safe limits, but otherwise there must be a minimum of interruptions due to power-equipment and primary-source failures.

Earlier British Post Office schemes have been powered by electronically controlled units feeding from one end only.[1] Manual changeover to stand-by units has been provided at the end feeding power and at the distant end, a method which has satisfactorily met the economic requirements of short schemes.

For the Clarenville–Sydney Mines link a new and more reliable design of equipment has been developed. An automatic no-break system provides an uninterrupted supply to repeaters unless equipments at both ends of the link simultaneously fail to deliver power.

The main improvement in the reliability of the equipment is the replacement of all high-power electron tubes by electromagnetic components. The automatic no-break system takes advantage of the fact that the rating of the repeater-isolating capacitors has been chosen to permit single-end feeding. Normally the link is fed from both ends, but in the event of one equipment failing to deliver power the link is powered from the other end without interruption to the cable supply. If the failure is due to a power-equipment fault, double-end feeding can rapidly be re-established by manual switching to the stand-by. During an a-c supply failure, single-end feeding must be maintained until the supply is restored. In view of the very reliable no-break a-c supply provided at both stations, the possibility of a simultaneous a-c supply failure at both ends of the link is extremely remote.

The power equipments at each end of the link must be capable of supplying the whole of the power to the cable, should one end fail, which requires that each should be capable of operating as a constant-current generator. If two constant-current generators are connected in series, unless precautions are taken, an unstable combination will result and the unit supplying the higher current will drive the other unit "off load." Manual adjustment could be provided to equalize the currents fed from the two ends, but a different solution has been developed in which one of the units is a constant-current master and controls the line current, while the other unit is a slave whose voltage-current characteristic in the normal operating range is such that its current is always equal to that of the master unit. If the slave unit fails, the master will take over the supply; if the master unit fails, the slave unit will take over the supply and automatically assume the role of a master unit. The first unit switched on to an unenergized link operates as a master generator and the other unit, on being switched on, automatically operates as a slave. Other than ensuring that the link is safe for energizing, there is no need for any co-operation between the two ends when putting the equipment into service.

Details of Method Employed

MASTER-SLAVE SYSTEM OF OPERATIONS

The output-current output-voltage characteristics for the equipments are shown in Fig. 1, and are the same for both regular and alternate equipments at both ends of the link. Any equipment can operate with any of the characteristics *a*, *b*, or *c*, Fig. 1.

Normally the choice between characteristics *a* and *b* is made automatically by the equipment. If the output voltage does not reach 80% (per cent) of the full link voltage (approximately 2 kv), the unit will have the slave characteristic *b* (*DCAB* in Fig. 1). If the output voltage reaches or exceeds 2 kv, the unit automatically switches to the master characteristic *a*, (*EAF* in Fig. 1). Once having switched to the master characteristic

---

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957.

J. F. P. THOMAS is with the British Post Office, London, England, and R. KELLY is with Standard Telephones and Cables, Ltd., London, England.

Vol.
Pt. 1
1956

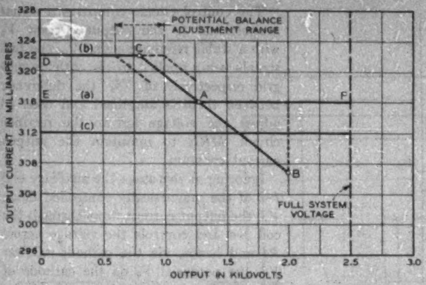


Fig. 1. Output current versus output-voltage characteristics. (a)—Master. (b)—Slave. (c)—Master versus master shutdown

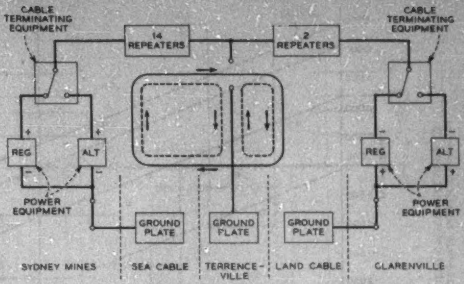


Fig. 2. Current paths on Clarenville–Sydney Mines link

the equipment does not automatically change back to the slave characteristic even when the output voltage falls below 2 kv.

When the link is energized, the first equipment switched to line will come on as a slave unit; its output voltage will then pass 2 kv and it will automatically be switched to the master characteristics and the complete link will be energized from one end. The second unit switched to line will come on as a slave unit and, having a higher output current, will drive down the output voltage of the master unit at the other end until the current of the slave unit has become equal to that of the master; see *DCA* in Fig. 1. The output voltage of this equipment will not exceed 2 kv and it will not switch to master. The crossover point of the two characteristics *A* will determine the potential fed from each end, and this can be adjusted as indicated by the dashed lines near *C*.

In an emergency, an equipment can be taken out of service by switching off the a-c supply, and the links will then be powered from the distant end only. Should the master end be switched off, the distant slave unit will switch to master characteristic as soon as its output voltage exceeds 2 kv. This abrupt disconnection of one equipment causes unnecessary voltage surges on the cable, and when an equipment is removed for normal maintenance purposes, the following procedure is adopted. The slave equipment is switched to the master characteristic by means of an external key for changing from master to slave or vice versa. (Its limitation is described in the next paragraph.) This leaves two master equipments feeding the cable, but any redistribution of voltage is slow since the currents are approximately equal. An external control (master/master shut-down), is then operated on the equipment to be taken out of service, changing its characteristic to that shown as *c* in Fig. 1. The output voltage of this unit will then be slowly reduced, and when it is zero the equipment can be switched off without causing surges on the cable.

The current deviations occurring over the slave characteristics, *c* in Fig. 1 (from +2% to −3%) are the maximum permitted by the tube design engineers for the tubes in the submerged repeaters, and are permitted only for short periods. The circuit associated with the manual switching of the equipment to the slave characteristic is therefore made inoperative if the output voltage exceeds 2 kv, since in this range the slave characteristic is the extension of the line *CAB* on curve *b* and the current is outside the permitted range. The slave characteristic over the range *CAB* is controlled by a voltage-sensitive circuit connected near the output of the equipment, and the stability of the characteristic against input voltage and component aging is the same as for the master characteristic. Changes in the distribution of the system potential due to supply variations and component aging are therefore small.

Over-all Current and Voltage Distribution

Facilities have been provided at the junction of the land and sea cables (Terrenceville) to connect a power ground to the center conductor of the cable. Normally this ground will be disconnected, but during installation it enables the land and sea sections to be energized separately and may subsequently be of assistance in the localization of cable faults near Terrenceville.

The full line in Fig. 2 shows the current path with normal double-end feeding, while the broken lines show the current paths when a ground is connected at Terrenceville.

The full line in Fig. 3 shows the voltage distribution along the link with normal double-end feeding, the broken lines *a* and *b* show the distribution with single-end feeding from Sydney Mines and Clarenville respectively, and the broken line *c* shows the distribution when a power ground is connected at Terrenceville.

Details of Equipments

Connection of Equipment to Cable

Each station is provided with two power equipments and one cable-terminating equipment interconnected as shown in Fig. 4. When the live side of the output of the regular equipment is connected to the cable via *SWD* and the link *LK1*, the alternate equipment output is connected to its own dummy load (equivalent of *RV1*) and vice versa.

The grounded sides of the regular and alternate equipments are common, and then connected via a removable link *LK2* to the sea ground. A safety resistor *R2* connects the sea ground to the station ground to restrict the rise in potential to 100 volts if the sea ground becomes disconnected. During maintenance on the sea-ground circuit the link *LK2* can connect the power-equipment ground to the station ground.

If, for maintenance purposes, it is necessary to feed from the distant end only, the link *LK1* can connect the cable to the power-equipment ground and, in addition, disconnect the live side of both of the power equipments from the line.

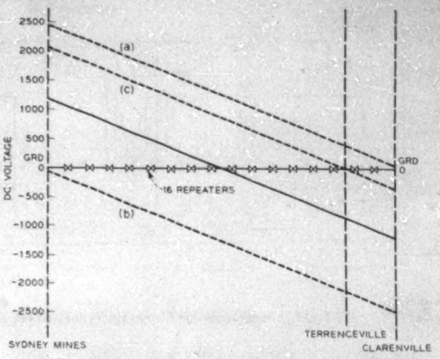


Fig. 3. Potential distribution on Clarenville–Sydney Mines link. (a)—Single-end feeding from Sydney Mines. (b)—Single-end feeding from Clarenville. (c)—Ground at Terrenceville; double-end feeding. (d)—Normal operation; double-end feeding

SAFETY INTERLOCKS

Safety interlock circuits are installed to protect the maintenance staff if the equipments are used incorrectly and are not the normal methods employed for controlling the power supplies. With double-end feeding, dangerous voltages are generated at both ends of the system, and when access is gained to any point in the equipments, personnel must be protected from the local and distant power sources.

To minimize interruptions to traffic, the units of the cable-terminating equipment are grouped under three headings, Fig. 4, namely:

1. Transmission equipment *S1* isolated from the d-c cable supply: this includes cable simulators and monitoring facilities not associated with the power supplies.
2. Equipment associated with the d-c supply that can be made safe without interrupting traffic.
3. Equipment that can be made safe only by interrupting traffic.

Group 1 is treated as normal terminal transmission equipment and is not interlocked. Groups 2 and 3, the power-changeover and cable-terminating boxes respectively, both require that the local power equipments be switched off before they can be made safe. The external panel-covers over both boxes have links disconnecting the a-c supplies to both of the local power equipments when the covers are removed, *LK3-LK6*. This will not interrupt traffic, since single-end feeding will continue from the distant station.

The doorknob switch of the power-changeover box automatically connects the live lead to ground at point *B*, when the door is opened. This will not interrupt traffic since the impedance of *L2* is high at carrier frequencies. Changing over the power equipments and grounding arrangements can therefore be performed without interrupting traffic.

The doorknob switch of the cable-terminating box automatically connects the center conductor of the cable to ground, at point *A*, when the door is opened. Traffic will therefore be interrupted if maintenance work is necessary within this box.

Access is gained to a power equipment by opening one or more of four doors. Each of the doorknob switches disconnects the a-c supply (*SWA*) and short-circuits the output of the equipment, *SWC*.

It will be appreciated that if the correct procedure is adopted the local power equipments will be switched off by *SWB* after the master/master shut-down procedure (described in the section "Master-Slave System of Operations," has been carried out, and not by removing the panel covers or opening the doors.

ELECTRICAL DETAILS OF POWER EQUIPMENT

Control Method

Fig. 5 is a functional simplified circuit diagram in which *TR1*, *MR1*, and *C1* represent a conventional unregulated power unit.

The control circuit compares a signal proportional to the output current V_0 with a stable reference signal V_R, the two signals being applied to the cathode and grid respectively of *V5;* the difference between them is amplified and used to adjust the voltage fed to the rectifier circuit *MR1*, to maintain the output current constant.

Ignoring at this stage the auxiliary coil *1–2* of the magnetically controlled diode *V3*, the output current flows through the coil *3–4* and controls the voltage across *R4* and hence, via the cathode-follower *V4*, the potential V_0 on the cathode of *V5*. V_R is a function of the output of a conventional electronic constant-voltage power unit (reference power unit) using a neon tube for its reference. The grid-cathode bias of *V5* controls the secondary impedance of the transductor *TD2* via the amplifier *V5*, *V6*, *TD1*. The gain of the control amplifier and the sign and magnitude of the normal impedance of the secondary of *TD2* are arranged to give a fall of approximately 0.25% in the current from full system load to short circuit.

Two advantages are obtained by employing a magnetically controlled diode for *V3* instead of an electrostatically controlled tube. The control is directly proportional to the output current, and is not dependent upon the stability of a series resistor, while the control circuits are isolated from the output-circuit voltage, which simplifies maintenance of the more complex parts of the equipment.

Current Characteristics

Without current flowing in the auxiliary coil *1-2*, on *V3* in Fig. 5, the equipment has a normal constant-current characteristic, the value of which, 322 ma (milliamperes), is preset by adjusting the mechanical position of the coil assembly along the main axis of *V3*.

Two neon tubes and two resistors form the bridge *V1*, *V2*, *R1*, and *R2*, which is balanced when the voltage drop across *R1* and *R2* equals the constant voltage across *V2* and *V1*, the output voltage at which this occurs being adjusted by *RV1*. At voltages below balance, current tries to flow from *y* to *x* and at voltages above balance it flows from *x* to *y*. The balance voltage corresponds to the voltage at which the slave-unit characteristic changes slope, *C* in Fig. 1. For voltages below balance the rectifier *MR2* prevents current from flowing in the winding *1-2* on *V3*, and in this range the constant current of 322 ma is maintained; see *DC*, Fig. 1. For voltages above balance, current flows in the winding *1-2* and

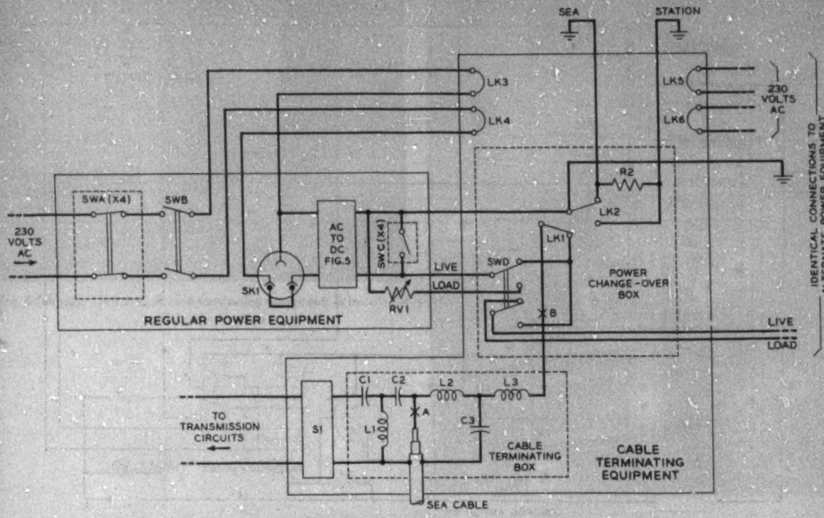


Fig. 4. Cable-terminating equipment and interconnections with power equipments

progressively decreases the output current *CAB*, Fig. 1. At 80% of the full link voltage, the contacts of relay *M* are operated and they disconnect the auxiliary coil *1-2* from the voltage-sensitive bridge and connect it across the reference supply. The current through *1-2* is then set by the fine current control *RV2* to make the output current the required 316 ma. As previously stated, relay *M* does not automatically switch back when the output voltage drops below 80 of full link voltage; the current characteristic of a master unit is *EAF* in Fig. 1.

The master/master shutdown characteristic, Fig. 1, curve *c*, is obtained by short-circuiting *R3*, which causes the constant current to fall to 312 ma. Adjusting *RV2* would be as effective, but short-circuiting *R3* does not permanently disturb the normal current setting.

Alarms

The equipment trips and gives both aural and visual alarms for +20% current +20% full link voltage and for the failure of certain auxiliary supplies that would damage the equipment.

Aural and visual alarms are provided for ±1.0% current and ±3% voltage, and for equipment changes from slave to master characteristic or vice versa. Visual indication is given if the a-c supply fails.

Reliability

Where possible, only components of proved integrity have been used. High-power thermionic tubes were excluded and the tube types employed were specially selected for long life. Electrolytic capacitors were excluded from all except one position, and in this case the component was divided into six units in parallel, the failure of all but one of these units causing only a slight increase in the output ripple.

Particular attention was given to the continuity of the output circuit. A failure of a power equipment for any reason other than an open-circuit in the output will not interrupt traffic, the system changing to single-end feeding from the far end. A disconnection anywhere in the d-c feed path will disconnect all power from the link. Relatively short-lived components in this part of the circuit were either duplicated in parallel or shunted by resistors capable of carrying the full line current.

Other Facilities

Each equipment has facilities for checking its over-all performance. A variable-ratio transformer can be introduced at *SK1*, Fig. 4, and, with the 4-position dummy load referred to earlier, the regulation against alternating input voltage and output load can be measured.

Provision is made for checking all the alarms, and the current can be measured at strategic points in the control circuit either when the equipment is normal or with the loop feedback disconnected.

Separate large-size meters are provided for measuring the output voltage and current to an accuracy of ±1.0%. A more accurate measurement of current is obtained from potentiometric measurements made across a standard resistor connected in series with the output.

ELECTRICAL DETAILS OF CABLE-TERMINATING EQUIPMENT

The majority of the electrical features of the cable-terminating equipment have already been considered.

Within the cable-terminating box, Fig. 4, are the power-separating filters; the high-pass filters *C1*, *C2*, and *L1* passing the carrier frequencies and the low-pass filter *L2* passing the direct current. The transmission equipment represented by the block *S1* is for convenience mounted in the cable-terminating cubicle. The extra low-pass filter *L3*, *C3* can be specially designed to prevent signals that are peculiar to the site (local radio stations, etc.), which are picked up in the power equipment, from being fed to the transmission circuits.

Metering facilities (not shown in Fig. 4) are provided to check the continuity

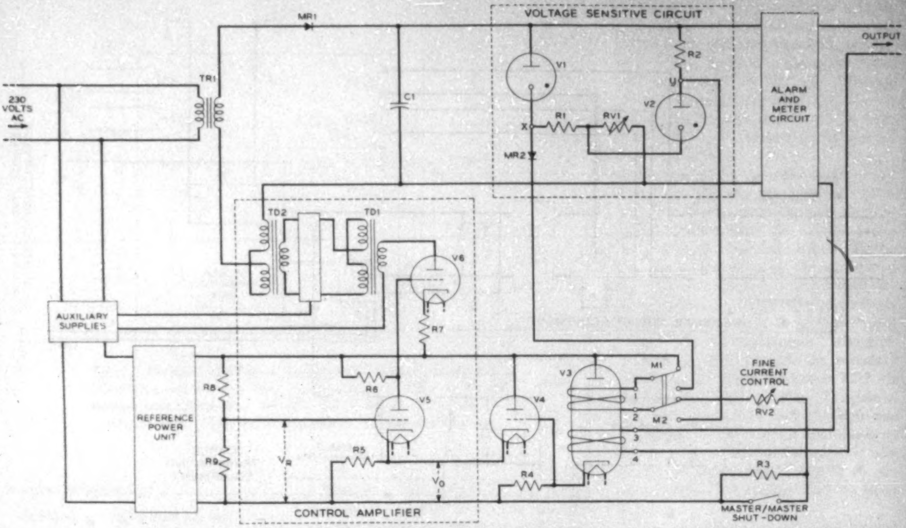


Fig. 5. Simplified schematic of power-feeding equipment

of the sea-ground circuit. A separate insulated wire is connected to the sea-ground plate and the continuity is checked by measuring the voltage drop along the ground cable. Aural and visual alarms are given if the sea ground becomes disconnected.

Current and Voltage Recorders

Current and voltage recorders are provided for both the regular and the alternate equipments, the values being measured at the points where the outputs of the power equipments enter the cable-terminating equipment.

A magnetic-amplifier unit drives the current recorder, the scale deflection being from −5% to +5% of the normal line current. The voltage recorder is connected across the ground end of a resistance potentiometer, the full-scale deflection being 3 kv.

The magnetic amplifier and potentiometer units are mounted in the cable-terminating equipment, but the four recorders and their associated supplies and alarms are mounted on a separate rack.

Mechanical Details

Fig. 6 shows two power equipments and a cable-terminating equipment as installed at each terminal station. The same cubicle frameworks are used for power and cable-terminating equipments, the power equipment consisting of two cubicles bolted side by side and fitted with doors at the front and back.

The top of the cable-terminating cubicle contains the power-changeover box and the center contains the cable-terminating box, while the transmission and ground-cable test circuits are located near the bottom. The main cable enters the bay at the top and passes behind the power-changeover box into the cable-terminating box. Access to the recorder units and cable simulators is from the rear.

Fig. 7 shows the front of one power equipment with the doors open. The left-hand cubicle contains the main high-voltage transformer, the electronic and magnetic parts of the control circuits, the reference power unit and the auxiliary supplies. The other cubicle contains the main rectifier and smoothing circuits, and the associated meter and alarm circuits.

Performance

Six power equipments and four cable-terminating equipments were manufactured for the link. Of these, four power equipments and two cable-terminating equipments were provided for the terminal stations and the remainder were for use on *HMTS Monarch* during the cable laying, and subsequently as off-station and training spares.

After individual testing, the power equipments were checked in pairs energizing a 16-section artificial cable constructed to simulate the Clarenville–Sydney Mines link. These tests were kept running for approximately 2 weeks on each equipment, and the line current was monitored with expanded-scale current recorders (from −1.0% to +1.0% of the normal current).

The equipments were completed early in 1956, and they have given satisfactory service both at the terminal stations and on board *HMTS Monarch*. From March to June 1956, all four units at the terminal stations were operating into their dummy loads and daily current and voltage readings were well within the required specifications. Since the laying of the sea section, the equipments have satisfactorily operated the completed link. When the link is in service, the working equipment at each end will be changed at 6-month periods and the changeover time will be staggered by 3 months each end. The equipment coming out of service will be immediately routine checked and adjusted if necessary.

The tests since installation confirm

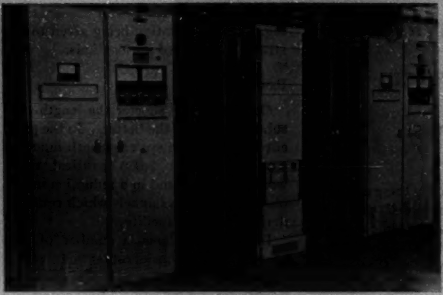


Fig. 6 (above). Power and cable-terminating equipments as installed at each terminal station

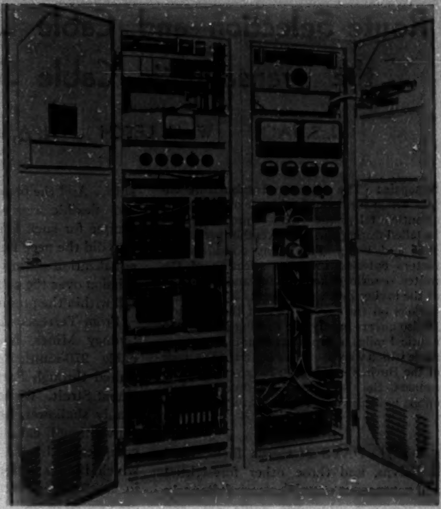


Fig. 7 (right). Power equipment

the laboratory and factory results that the regulation is better than ±1 ma for any alternating input voltage from 195 to 265 volts in the frequency range of 40 to 70 cycles for output voltages of 0 to 3 kv. The relatively long correction period of the magnetic-amplifier control circuit (about 0.5 second) is satisfactory with the type of no-break supply installed at the stations.

Another useful by-product of double-end feeding is that, when there is a shunt fault on the system, the voltages supplied from the two terminal stations give an indication of the fault position. Many factors will control the accuracy of the location, e.g., magnitude and position of the fault, and how nearly the currents fed from the two ends are equal. Calculations, confirmed by tests made with shunt faults introduced at Terrenceville, show that any continuous shunt fault that will affect transmission (to the extent of operating the 1-db pilot alarms) will be detected to an accuracy of ±1.0%. The limit is set by the accuracy with which the link voltage distribution can be measured.

Conclusions

The power-feeding system described is suitable for submerged-repeater schemes that can, in an emergency, be temporarily powered from one end only. For locations within reasonable reach of a central catastrophe-spare store the equipment should be sufficiently reliable not to need duplication at both terminal stations. For future schemes this will provide a method which is economically attractive compared with the present single-ended methods and which offers numerous electrical advantages.

The routine maintenance required on the power equipment could be further reduced if the few remaining electron tubes were replaced by electromagnetic components. On schemes where short interruptions to traffic do not involve a relatively high loss of revenue it would then be unnecessary for the local staff to maintain the high-voltage equipment, and the expensive no-break a-c supplies could be abandoned. The latter depends upon the probability of the primary sources at the terminal stations failing simultaneously.

Acknowledgments

The authors wish to express their thanks to many of their colleagues in the British Post Office and Standard Telephones and Cables, Ltd., who contributed to the developments described in the paper. The permission of the engineer-in-chief of the Post Office, and of Standard Telephones and Cables, Ltd., to make use of the information contained in the paper is also gratefully acknowledged. The equipments for the Clarenville–Sydney Mines link were manufactured by Standard Telephones and Cables, Ltd., at North Woolwich, London.

Reference

1. The British Post Office Standard Submerged-Repeater System for Shallow-Water Cables (with Special Reference to the England-Netherlands System), D. C. Walker, J. F. P. Thomas. *Proceedings,* Institution of Electrical Engineers, London, England, vol. 101, pt. I, Feb. 1954, p. 190.

Route Selection and Cable Laying for the Transatlantic Cable System

J. S. JACK W. H. LEECH H. A. LEWIS

Synopsis: The repeatered submarine cables which form the backbone of the transatlantic telephone cable project were installed during the good-weather periods of 1955 and 1956. This paper considers the factors entering into the selection of the routes, describes the planning and execution of the laying task and presents a few observations on the human side of the venture. It also covers briefly the routing of some 55 nautical miles of repeatered submarine-type cable which were trenched in across the neck of the Burin Peninsula in Newfoundland to connect the Terrenceville submarine terminus with the cable station at Clarenville.

IN THE DAYS of Cyrus Field, Lord Kelvin, and those other foresighted and courageous entrepreneurs of the early transoceanic submarine cable era, the risks involved in selecting a route and laying such a cable must have appeared formidable beyond description. And indeed they were, for not until the third attempt was a cable successfully laid.

Today the hazards may be somewhat more predictable, our knowledge of the ocean bottom more refined, and our tools improved, but only to a degree. The task still remains extremely exacting in its demands for sound engineering judgment, careful preparation, high-grade seamanship, and good luck, weatherwise. For the basic methods now in use are still remarkably like those employed on the *Great Eastern* and other early cable ships, and the meteorological, geographical, and topographical problems have changed not at all.

In the current transatlantic project, the first transoceanic telephone cable system, there are two submarine links. Between Clarenville in Newfoundland, and Oban in Scotland there lie some 1,850 nautical miles* of North Atlantic water, most of it deep and all of it subject to weather of unpredictable and frequently unpleasant nature. The bridging of this required the laying of two 1-way cables over carefully selected routes, using an available cable ship. And the presence in these cables of 102 flexible repeaters posed problems unique for such long and deep cables, as also did the need for trimming the system equalization during laying so that transmission over the completed system would fall within the prescribed limits.

From Terrenceville, Newfoundland, to Sydney Mines, Nova Scotia, a single-cable 270-nautical-mile path was required through Fortune Bay and across Cabot Strait. While this water is considerably shallower, here again a relatively conventional cable-laying problem was complicated by the presence of repeaters which, in this section, were rigid units, 14 in number. (Two additional repeaters are located in the 55-nautical-mile section which is trenched in across Newfoundland from Clarenville to Terrenceville.) Trimming of system equalization was also required.

These cables were laid during the spring and summer months of 1955 and 1956. And the preparation for the laying required many months of effort in fields which were for the most part quite foreign to the usual scope of land wire telephone activity. Some appreciation of the problems encountered in this phase of the venture may be gained from the following sections.

North Atlantic Link

Route Selection

The first successful transatlantic telegraph cable was laid across the North Atlantic in 1866. Since that date 15 direct cables have been laid and five by way of the Azores. The approximate routes of these cables are shown in Fig. 1. It is at once evident that the shortest and possibly the best routes were already occupied so that selection of routes for the two transatlantic telephone cables could be expected to present some difficulty.

Some of the more important considerations which guided the selection were: 1. route length, 2. clearance for repairs, 3. trawler and anchor damage possibilities, and 4. terminal locations suitable for repeater stations, with staffing in mind, as well as facilities for onward routing, due consideration being given to the strategic aspects of the locations.

Route Length

Obviously, the shorter the length of a submarine route, the better. In the present instance, any system length much in excess of about 2,000 nautical miles would have resulted in a reduction in the number of voice channels which could be derived from the facility.

In Fig. 2 is shown a number of the routes to which consideration was given in the early planning stages. The distances shown are actual cable lengths and include an allowance for the slack necessary to assure conformance of the cable to the profile of the ocean bottom.

Route 1, from Eire to Newfoundland, at 1,770 miles, is the shortest route and in point of fact was provisionally suggested in 1930 for a new cable. But the difficulty of onward transmission of traffic to London made this route unattractive.

Route 2, from Newfoundland to Scotland, compared favorably in length with route 1, but its adoption was dependent upon location of a suitable landing site in Scotland.

Route 3, from Newfoundland to Cornwall, England, approximated 2,000 miles laid length and would have been very attractive had not so many existing cables terminated in southern Ireland or the southwest corner of Cornwall, which would lead to a great amount of congestion and consequent hazards to the telephone cables.

Route 4, from New York to Cornwall, was too long to be considered, as its length amounted to some 3,200 miles.

Routes 5, 6, 7, and 8 were indirect via the Azores. They were attractive, as only relatively short lengths were involved and suitable sites for intermediate cable stations could have been found on one of the several islands in the Azores. But difficulties attendant upon landing rights, and staffing problems in foreign territory were foreseen.

Clearance for Repairs

Repair of a faulty cable or repeater necessitates grappling, and in deep water this is likely to be a difficult operation. To avoid imperiling other cables while grappling for the telephone cables and, conversely, to provide assurance against accidental damage to the telephone cables from the grappling operations of others, it was considered essential that the route selected provide adequate clearance from existing cables. Suitable clearance is considered to be 15 to 20 miles in the

---

A Conference Paper presented at the Symposium on the Transatlantic Submarine Telephone Cable, during the AIEE Winter General Meeting, New York, N. Y., January 21–25, 1957.

J. S. Jack is with the American Telephone and Telegraph Company, Long Lines Department, New York, N. Y.; Capt. W. H. Leech is with the British Post Office, London, England; H. A. Lewis is with the Bell Telephone Laboratories, Murray Hill, N. J.

* A nautical mile, as used herein, is 6,087 feet, 15.3 per cent longer than a statute mile.

Vol. Pt. 1 1956

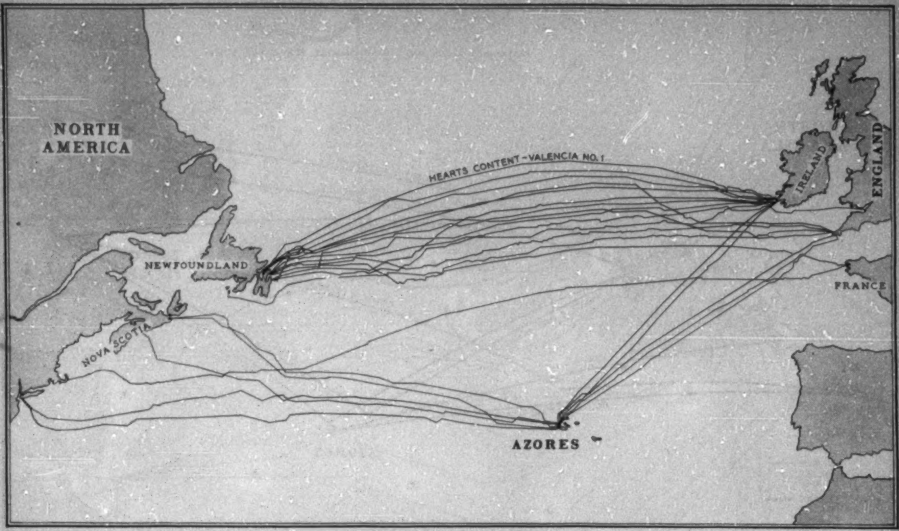


Fig. 1. Telegraph cables in the North Atlantic

ocean, with less permissible in the shallower waters of the continental shelves.

TRAWLER AND ANCHOR DAMAGE POSSIBILITIES

It is probable that fishing trawlers cause more interruption of submarine cables than any other outside agency. Cables laid across good fishing grounds are always liable to damage from fouling by the otter boards of the trawlers. Final splices, either initial or as a result of repair operations, are especially vulnerable to damage because of the difficulty in avoiding slack bights at such points. It was desired therefore to avoid fishing grounds if at all possible.

If cables are laid in or near harbors frequented by merchant shipping, damage must be expected from vessels anchoring off shore in depths of less than 30 fathoms and proposed routes should therefore avoid such areas.

CABLE TERMINAL SITING

Location of the cable terminal stations must be considered from the standpoints of suitability of shore line for bringing the cables out of the water and also from the standpoint of amenities for the staff. This latter factor is most important in keeping a permanent well-trained staff. For example, owing to staff difficulties, it was necessary to move a terminal station of one company from the west side of Conception Bay in Newfoundland to a site within easy reach of St. Johns. A further factor in proper siting of the cable terminals is consideration for onward routing of the circuits carried by the cables.

And finally, in view of the importance, generally, of submarine cable facilities, it is considered desirable to avoid cable terminal locations in or near a potential military target area and, if possible, consideration should be given to underground protective construction for the terminal stations.

PRELIMINARY SELECTION

The routes for the telephone cables were considered in the light of the foregoing, and after preliminary discussion it was agreed that the two new cables should lie north of all existing cables, should avoid ships' anchorages, and should lie on the best bottom which could be picked, clear of all known trawling areas.

In 1930, American Telephone and Telegraph Company, in conjunction with the British Post Office, gave serious consideration to the laying of a single coaxial telephone cable between Newfoundland and Frenchport, Ireland (route 1, Fig. 2). A tentative route was plotted and the cable ship *Dominia* steamed over this, taking a series of soundings. These soundings indicated that good bottom was to be found about 20 miles north of the Hearts Content–Valencia no. 1 cable of 1873. This cable was the most northerly of the telegraph cables spanning the Atlantic. Study of its life history indicated that faults clear of the continental shelf were few and far between throughout its long life.

The latest British Admiralty charts and bathymetric charts of the U. S. Hydrographic Office for the North Atlantic Ocean were scrutinized and, from these and a study of all other relevant data, two routes were plotted which appeared to fulfill the necessary requirements so far as possible. However, it was agreed that if possible the selected routes should be surveyed so that minor adjustments could be made when desirable.

LANDING SITES

East End

It was now necessary to find suitable landing sites, with regard for the decision that the telephone cables should be routed north of all other existing cables.

On the British side it was necessary to look north of Ireland. The North Channel, the northern entrance to the Irish Sea, divides northern Ireland from Scotland and had this channel been suitable, the telephone cables might have been run

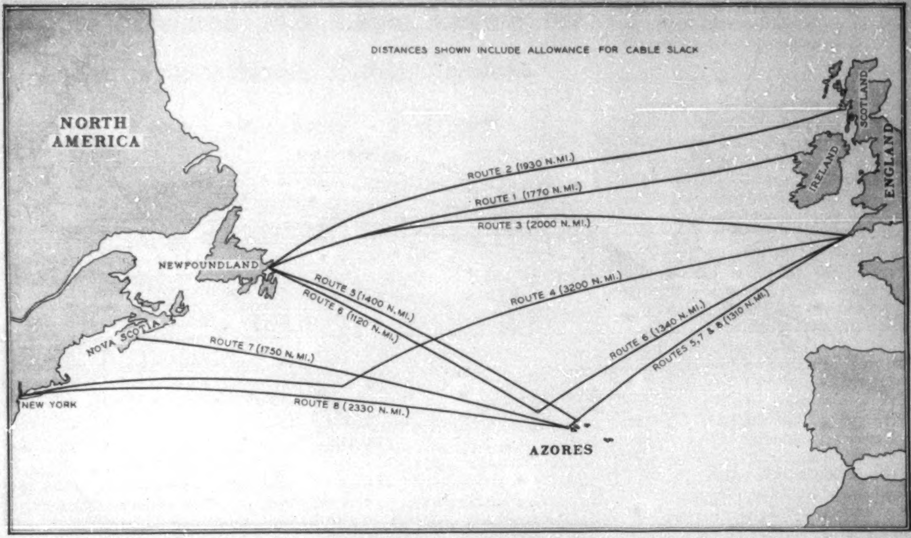


Fig. 2. Tentative telephone cable routes

through it to a terminal station on the southwest coast of Scotland in the vicinity of Cairn Ryan. However, the tidal streams through the channel are strong, at least 4 to 5 knots at spring tides; the bottom is rocky and uneven, with overfalls, and any cable laid through it would have a very short life indeed.

It was therefore necessary to search farther north. The west coast of Scotland presents a practically continuous series of deep indentations and bald, rocky cliffs and headlands. The chain of the Hebrides Islands stretches almost uninterruptedly parallel with and at short distances from the coast. It was obviously most desirable to land the cable on the Scottish mainland, and close to rail and road communication if possible.

From previous cable maintenance experience it was known that the Firth of Lorne which separates the island of Mull from the mainland was a quiet channel, little used by shipping or frequented by trawlers and with tidal streams which were not strong. Earlier passages of Post Office cable ships through the Firth had yielded a series of echo sounding surveys which indicated that, except for a distance of about 5 or 6 miles in the vicinity of the Isles of the Sea, the bottom was fairly regular. Several small bays on the mainland side of the Firth just south of Oban appeared from seaward to be very suitable landing sites and this was confirmed by a survey party, which selected a small bay locally named Port Lathaich for the cable landing and site of the station.

The fore shore was mainly firm sand with outcroppings of rock which could be avoided easily when landing the cables. The seaward approach was clear of danger and there was ample room to land two cables with a separation on the shore of some 30 yards.

Port Lathaich is only about 3 miles by road from Oban. Additional land cables would be necessary, however, to carry traffic to the main trunk network. From a strategic point of view, although Oban might possibly be considered a target area, the cable landing was sufficiently remote to be relatively safe, especially if the cable terminal station was sited in the rocky hillside. To ascertain whether any serious chafing or corrosion would result if cables were laid over the uneven bottom in the Firth, some 8 miles of coaxial cable with E-type armoring were laid over the area and recovered after 2 years. There was no evidence of any chafing or corrosion. It was therefore decided that the telephone cables should be routed through the Firth of Lorne to the cable terminal station site at Port Lathaich.

West End

The choice of a suitable cable landing in Newfoundland was more difficult to make, in view of the rugged and sparsely populated nature of the country. From Fig. 3 it will be seen that the existing telegraph cables spanning the Atlantic land either just north of St. Johns, in Conception Bay, or in Trinity Bay. North of Cape Bonavista the coast becomes more broken, and the sea approach is not good. Accordingly, there was no good alternative to routing both telephone cables into Trinity Bay, close to and northwest of the telegraph cable landing at Hearts Content on the southern shore of the bay. A survey party made an extensive examination of all likely places on the western side of the bay from Cape Bonavista in the north to Bull Arm at the southern end where, incidentally, the first successful telegraph cable was landed. Careful consideration of all of the places visited led to the agreement that Clarenville was the best site for a landing and for a cable terminal station.

Clarenville is at the head of the Northwest Arm of Random Sound. It is a junction on the main railway, and a good road to St. Johns will pass through the town in traversing its course from St. Johns to Port aux Basques. Clarenville has a growing population of some 1,600 inhabi-

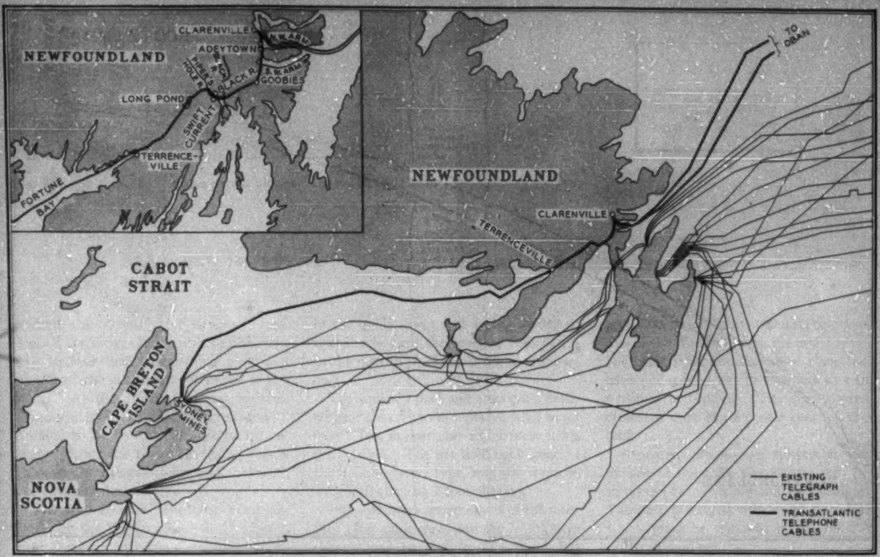


Fig. 3. Cable landings on Newfoundland, and final route of overland section of the Clarenville–Sydney Mines cable

tants, with stores and repair facilities of various sorts. Good cable landing sites are available just out of town and the approach from the sea up the Northwest Arm presents no navigational difficulties. Such few small vessels as ply to Clarenville during the summer months are not likely to interfere with the cables.

Final Route Agreement

Having agreed upon Clarenville, Newfoundland, and Oban (Port Lathaich), Scotland, for shore terminations, it was possible to complete the routes for the two cables as shown in Figs. 3 and 4. The final routes are clear of existing cables and they avoid crossing known trawling areas and anchorages. The cable stations are well sited with regard to staff amenities, accessibility, and strategic requirements. Soundings taken during the laying of the two cables showed a very even bottom except in the Firth of Lorne and one or two places in Trinity Bay. The general profile of the route is shown in Fig. 5.

It is considered that these routes have been selected with care and that they meet all the requirements of a well-planned cable project. Time alone will tell how well the objectives have been met.

Cable Laying

Early Methods

In 1865, when the legendary *Great Eastern* was pressed into service to lay the first successful transoceanic telegraph cable, she was fitted out with certain special cable-handling gear. The need for such gear had been amply demonstrated by events which transpired during two earlier and unsuccessful attempts by *HMS Agamemnon* and *USS Niagara*.

For her assignment, the *Great Eastern* was fitted with three large tanks into which her cargo of cable could be coiled. She was also provided with a large drum about which the cable could be wrapped in the course of its passage from the tanks to the sea. This drum was connected to an adjustable braking mechanism which provided the drag necessary to assure that the cable payout rate was correct with relation to the speed of the vessel. In addition, a dynamometer was provided so that the stress in the cable would be known at all times. A large sheave fitted to the stern of the ship provided the point of departure of the cable in its journey to the sea bottom.

On Friday, July 13, 1866, The *Great Eastern* steamed away from Valencia, Ireland, and 14 days later, on July 27, she arrived off Trinity Bay, Newfoundland, and completed the landing of the western shore end.

HMTS Monarch

Early in the planning for the transatlantic project it was realized that in no small measure the success of the venture would depend on availability of a vessel suitable for laying the cables. It was extremely fortunate that one of the partners to the enterprise was also the owner and operator of the largest cable ship in all the world, and one well suited to the task at hand.

The twin-screw cable ship *Monarch*, Fig. 6, was built for H. M. Post Master General by Messrs. Swan, Hunter, and Wigham Richardson, Ltd., at their Neptune Works, Walker-on-Tyne. She was completed in 1946. This ship is of the shelter deck type, having principal dimensions as follows:

Length over-all.......	482 feet 9 inches
Breadth molded......	55 feet 6 inches
Depth molded to shelter deck...........	40 feet 0 inches
Gross tonnage........	8,056

The ship has an overhanging bow which carries three cable sheaves, a cruiser

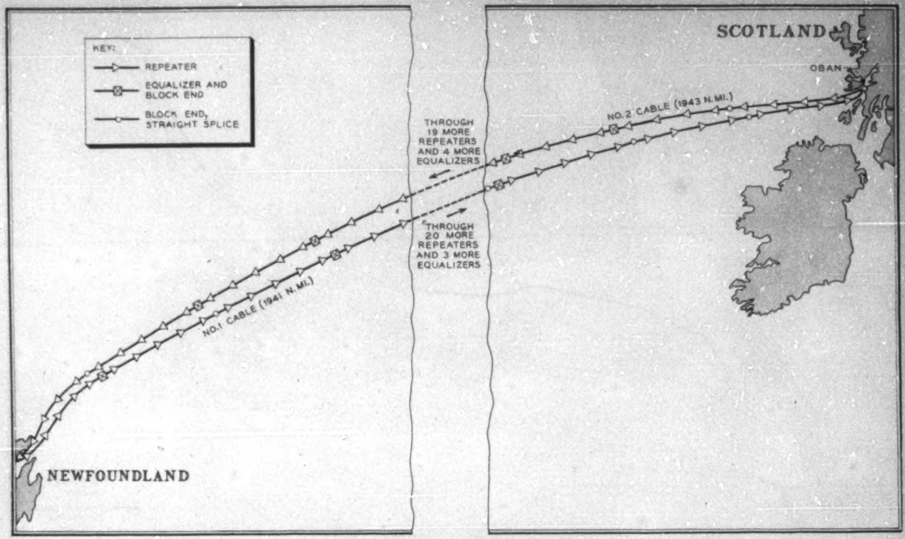

KEY:
REPEATER
EQUALIZER AND BLOCK END
BLOCK END, STRAIGHT SPLICE
SCOTLAND
OBAN
NO. 2 CABLE (1943 N. MI.)
THROUGH 19 MORE REPEATERS AND 4 MORE EQUALIZERS
THROUGH 20 MORE REPEATERS AND 3 MORE EQUALIZERS
NO. 1 CABLE (1941 N. MI.)
NEWFOUNDLAND

Fig. 4. Transatlantic telephone cable routes

stern with the after paying-out cable sheave offset on the port quarter, a semibalanced rudder having extra large surface, and a cellular double bottom extending from the collision to the aft peak bulkheads. Both main and shelter decks are steel and they extend her complete length.

The cable is carried in four welded steel cable tanks fixed to the tank top plating. These are arranged along the ship's center line in a fore-and-aft direction forward of the main propelling machinery space. They are each 41 feet in diameter and have the following cubic capacities:

Tank, No.	Coiling Space	Gross Cubic Feet
1	33,730	40,170
2	31,820	38,460
3	30,865	37,375
4	30,230	36,300

The opening in the shelter deck above each tank is a circular hatch 8 feet in diameter.

A watertight cone of steel plates is built in the center of each tank to ensure against fouling of the cable during payout. Further control of the cable is provided by a crinoline, Fig. 7, which is a circular spider of steel tubing normally suspended from 1 foot to 3 feet above the top layer of cable in the tank. The crinoline tends to prevent flying bights of cable and also provides a safety platform, in case of trouble for the men who work in the cable tanks. Each crinoline may be raised and lowered by an electric motor drive.

The maximum cable-carrying capacity is approximately 5,000 long tons, or almost 2,000 miles of the deep-sea type of cable used on this project if no repeaters were involved.

The *Monarch* is driven by two steam engines. The maximum propeller revolutions are estimated at 110 per minute, giving a ship's speed of about 14 knots.

Two cable engines are fitted forward, both capable of being used for picking up or paying out. These are driven by electric motors having a maximum rating of 160 horsepower which will permit picking up at a rate of 0.9 nautical mile per hour with a stress of 20 tons, or at 3.5 knots with a stress of 5.3 tons. The drive system is constant current, so designed that a uniform torque may be held at the drum for any setting of the speed control. When paying out, these motors operate as generators to provide electrical braking, and auxiliary mechanical brakes are also provided.

A single cable engine is fitted aft and this is the main paying out gear. In addition to the electric brake, the aft engine is provided with a multiple drum externally contracting band brake, manually adjustable and water-cooled, and with a further auxiliary fan brake. The fan shaft is driven in such a manner that when cable is being paid out at approximately $8^3/_4$ knots, the fan will revolve at 1,000 rpm and absorb 120 brake horsepower. Adjustments in this are effected by varying the amount of opening in the fan shroud so that as little as 27 brake horsepower may be absorbed.

Dynamometers, both fore and aft, provide for measurement of the cable tension. Taut-wire gear is furnished on the starboard quarter to provide an effective means for calculating the amount of slack paid out. With this gear, steel piano wire, anchored to the bottom, is paid out at constant tension and provides a rough measure of distance steamed over the ground. A test room with trunks to each cable tank is provided on the shelter deck and fitted with instruments for measuring and locating faults.

Modifications for Flexible Repeaters

In the normal cable-paying-out process, the cable is drawn from the tank, carried along fairleads to the holdback gear (a mechanism for applying slight tension to the cable so that it will snub tightly

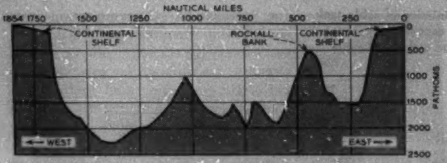


Fig. 5 (above). Profile of ocean depths between Clarenville and Oban

Fig. 6 (right). HMTS Monarch

around the drum), and then wrapped around the drum of the cable engine from two to four turns, depending upon the weight of the cable and the depth of the water. At the drum, a fleeting knife is fitted which pushes over the turns already present to make way for the oncoming turn. From the drum the cable passes through the dynamometer and thence to the overboarding sheave.

The Bell System repeaters, manufactured by the Western Electric Company, were designed with the objective of making them act as much like cable as possible.[1] Despite this, their presence introduced a loading and laying problem as their ability to bend without injury is limited to about a $3\frac{1}{2}$-foot radius, and their structure is such that unnecessary bending may present a hazard to their watertightness. As the majority of the sheaves and drums of the conventional laying gear are considerably smaller than 7 feet in diameter, a number of modifications were required in the *Monarch's* equipment to satisfy the repeaters.

For the most part, the new gear was designed by the Telegraph Construction and Maintenance Company, Ltd., to broad requirements supplied by the American Telephone and Telegraph Company. The modifications included providing the port bow sheave with a flat tread to bring its diameter to 6 feet 10 inches, and replacing both forward dynamometers and the aft dynamometer by a new design employing a 7-foot wheel in a pivoted A-frame bearing on Elliott pressure-type load cells. In this arrangement port and starboard forward drums were replaced with drums of maximum possible diameter without a major change in the complete gear. This diameter proved to be 6 feet 10 inches on the tread. The after paying out drum was replaced with one having a 7-foot diameter. The forward port and aft cable drums were equipped with ironing boards. (An ironing board is a curved shoe placed adjacent to the cable drum and spring loaded so that it will force the repeater to conform to the curvature of the drum as it passes on.)

The forward port and starboard draw-off gear sheaves were replaced with larger ones 7 feet in diameter which were made traversable. The aft hold-back gear, of the double-sheave type, was also replaced with units having 7-foot sheaves.

Fig. 8 shows a general view of the modified after cable gear, and the 7-foot stern sheave may be seen in Fig. 9. A line schematic of the gear is shown in Fig. 10.

Roller-type fairleads shaped into arcs of minimum $3\frac{1}{2}$-foot radius were fitted at each cable tank hatch, with smaller roller guides at convenient points to assure fair lead of the cable from the tanks to the cable machinery. Electric hoisting gear was provided for the lower crinoline in each tank as it was necessary to raise the crinoline whenever a repeater left the tank.

The test room was greatly enlarged and fitted with the special gear necessary for powering and measuring the system during laying.

LOADING CONSIDERATIONS

When the ship is loaded, the cable is coiled carefully in the tanks, layer upon layer, each layer being called a "flake." The coiling is started from the outside of the tank and progresses clockwise toward the center so that the armor is untwisted one revolution for each complete turn in the tank. When paid out in the reverse order, this twist is restored.

Handling of the repeaters during loading presents a problem because of the need to restrict their bending. After some experimental work, splints were devised to provide the needed rigidity. These consisted of two angle irons each 12 feet long and equipped at the ends with cold-rolled steel rods ranging in length from $1\frac{1}{2}$ to 6 feet. By this device it was possible to maintain rigidity over the main central portion of the repeater, including the junction of the core tube with the end nosing, and provide limited flexibility along the outer ends of the core tubes which are less sensitive to bending. The splints were removed, once the repeater reached the tank.

Repeaters are always stowed at the outside of the flake where they need be subjected to only a minimum of bending. They are protected with wood dunnage, which must be removed before the repeater is paid out.

With these modifications all repeaters and equalizers were laid successfully from either forward or aft gear at a cable speed of around 3 knots.

TESTING AND EQUALIZATION

Purpose

Once a submarine cable system has been installed, it is accessible only at the ends for adjustment to improve performance, save at great difficulty and large cost. As some irregularities cannot be corrected from the ends, it behooves the designers to discover and account for such irregularities and to correct them before the cable is finally on the ocean bed.

The laying period offers the last opportunity for accomplishing this and, indeed, all too frequently, also the first. This fact, coupled with the broadband design of the link and the presence in the system of active elements (the repeaters), necessitated a very comprehensive program of tests and measurements during laying.

The program had three specific purposes: 1. to detect immediately any fault which might develop during laying, 2. to permit the design and execution of corrective system adjustments while en route so that transmission performance of the completed link would fall within specified objectives, and 3. to gather data on system characteristics at intermediate points for eventual use in fault location or

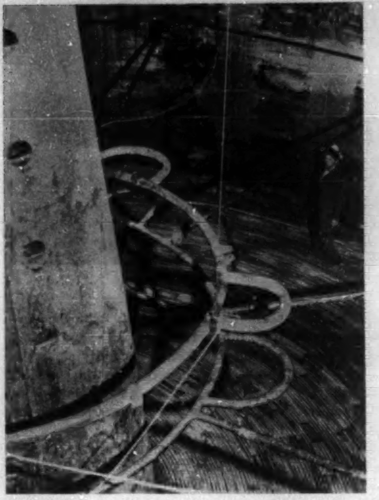


Fig. 7. Interior of cable tank showing central cone, crinoline, and a flake of cable

in aging studies. The need for parts 1 and 3 of the program is more or less self evident, but part 2 merits some further discussion.

In an ideal submarine cable system in an average environment, the attenuation of the cable from one repeater to the next would be offset exactly across the frequency band by the gain of the following repeater. Such a result is never achieved in practice, as the temperature and pressure environments (which affect cable attenuation) cannot be known precisely in advance, and the cable structure itself cannot be manufactured for mile after mile without variation in transmission characteristic. Additionally, the mechanics of the laying process induce minor changes in the physical structure of the cable which are reflected in attenuation changes.

If such deviations from desired characteristic produced only differences from the specified system gain objective, compensation could be readily applied at the ends of the submarine link. Unfortunately, this is only partly the case. Their more important effect is the resulting misalignment of operating levels of individual repeaters from design objective.

Misalignment magnitudes must be watched carefully for, at best, misalignment narrows the system latitude for seasonal temperature changes and for aging, and at worst it can result in intolerable system noise. If a repeater is preceded by too much cable the signal-to-noise ratio at the repeater input will be less than desired because of thermal noise. In the opposite case of too little cable, the signal level will be too high and the resulting overloading in the repeater will also affect the signal/noise adversely. Once present on the signal, the noise cannot be removed, and so the cure for excessive misalignment must be applied before the misalignment has developed. Adjustments at intermediate points along the route must therefore be contemplated.

Testing Program

The program which was evolved to meet the three objectives outlined was meticulously reviewed and practiced before the start of laying, and various forms were prepared for entering data and plotting and evaluating results. This was essential to avoid wasting effort or missing valuable data. The wisdom of this was fully apparent to all involved after experience with the close time schedules and the mental tensions which developed during the actual laying.

Staffing for testing was provided by crews of two or three trained engineers located at the transmitting cable station and on shipboard. Those on the ship served $4^1/_2$-hour watches at 9-hour intervals, which permitted a reasonable amount of rest, and avoided continuous "dog watch" duty by any one crew.

Close contact between shipboard and cable station crews was essential, and was achieved by cable and radio order circuits (or "speakers"). Communication from shore to ship when the cable was powered made use of the standard cable order wire circuit at the cable station to apply a signal in the frequency band 16 to 20 kc. The signal was demodulated and amplified aboard ship by a special stripped version of this same gear. The radio order circuits employed special land antennae and equipment and, for the most part, the ship's standard single-sideband telephone set, although other equipment at medium frequency was sometimes used for short distances. Radio telegraph with hand keying was available for backup when conditions were too poor for the radiophone sets.

Plans called for powering the cable at all times except when splices were being made. This was necessary, of course, for the measurement program, but it also provided additional assurance of the safe laying of repeaters, as the glassware and tungsten heaters of the electron tubes are more resistant to damage when hot. Power for the first half of each crossing was provided from the cable station. Beyond this point, the required voltage would have become excessive and so the shipboard supply was inserted into the series power loop and its voltage adjusted in proportion to the amount of second-half cable actually in the loop.

Monitoring against the possibility of faults was accomplished by measurement of a pilot tone at 160 kc, transmitted over the cable at all times except when data were being taken or power was turned down. Audibly alarmed limits were set on the measurement to indicate any significant deviation in transmission. In actuality, all unanticipated alarms received were found to have resulted from frequency or voltage shifts in the primary shipboard supply for the measuring equipment.

During the design of the system,[2] consideration of the misalignment problem had indicated the desirability of splitting the cable for each crossing into a number of sections, called "ocean blocks." These contained either four or five repeaters, and were 150 to 200 miles in length. In loading the ship, the two ends of each ocean block were left accessible for connection to the test room and for splicing operations.

Measurements made in the spring of 1955 off Gibraltar had indicated an unexpected change in attenuation called "laying effect,"[3] which required some last-minute adjustment of the repeater section lengths. With incorporation of these changes, it was known that the factory lengths of cable between repeaters

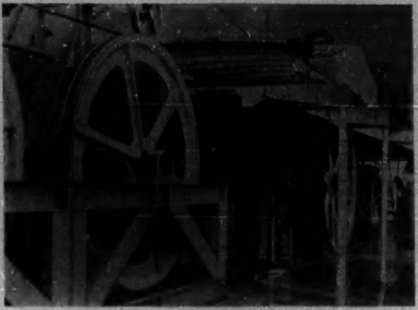


Fig. 8. General view of modified after cable gear (one of two hold-back sheaves, drum with fleeting knife and ironing board and, at extreme right, dynamometer sheave)

Fig. 9. Cable payout over the stern

were adequate to keep misalignment within an ocean block within reasonable limits. The system could then be equalized between ocean blocks so that the signal level at the first repeater of a new block would be approximately correct, and the total system noise thus would fall within limits.

This equalization was accomplished in two ways. Excess cable of the order of 1/2 mile to 3 miles in length was provided at the top end (first end out of the cable tank) of each ocean block. Based on measurements, this could be cut longer or shorter than the nominal spacing of repeaters, so that the repeater gains and cable losses would be matched at some frequency in the band. Residual deviations in other parts of the band could then be mopped up if necessary by inserting a simple equalizer, housed in a container similar to those used for the repeaters. Ten such equalization points were provided in each cable.

In practice, sending levels were adjusted at the cable station to give test tones at the grids of the output tubes of the repeaters which, if the system equalization were perfect, would be flat across the frequency band, and at the proper level. These tones were measured on shipboard at the end of the ocean block being paid out. The results were plotted against mileage, with one sheet for each frequency being measured. Because of the laying effect and of temperature and pressure changes on the cable as it progressed to the bottom, these plots displayed a slope.

The value of loss (or gain) to be ascertained for each frequency was that which would exist when the entire ocean block was on the bottom. To obtain this, it was necessary to extrapolate the curves to the mileage point representing the end of the block in question. The extrapolation was required to avoid stopping the ship at the end of the block, and so had to anticipate the time needed for turning over and cutting the cable end at the proper point, and making one or two splices (depending on whether or not an equalizer was inserted at the point in question).

Having read the extrapolated values from the curves, these were compared with objectives for the particular block junction, and the deviations plotted. Transparent overlays, showing the net effect of each of several types of equalizer combined with varying amounts of cable around the nominal spacing, greatly facilitated the final decision as to cutting point and equalizer choice. This implementation of the system's undersea equalization represented a very large part of the effort required of the testing crews during laying.

Additional data gathered for fault location, aging studies, and for other purposes included precise determination of repeater crystal fequencies[2] on the bottom, gain frequency runs to show up any fine-grained structure which might exist in the band, and values of the current and driving voltage.

Copper resistance and capacitance measurements proved of dubious value; in the first case because of the temperature-resistance characteristic of the electron tube heaters; in the case of capacitance, probably because of polarization effects in the castor oil capacitors used in the repeaters.

Shipboard Test Equipment

A new test room had been equipped for making these measurements with transmitting and receiving transmission measuring sets,[3] including the crystal test panels. These sets were provided in duplicate to forestall difficulty should one set develop trouble during the laying. The transmitting consoles were required only for use in calibrating the receiving sets, and for some measurements which were made on individual ocean blocks in the ship's tanks.

Additional gear in the test room included a cable current power supply,[4] and a "Lookator," which is a pulse echo type of fault locator useful from a point in the cable to the adjacent repeater on either side.

LAYING SEQUENCE

HMTS Monarch is the largest cable ship afloat, with capacity for about 2,000 miles of the type-*D* deep-sea cable in her tanks. However, because of the inherent limitation on their bending radius, the presence of flexible repeaters in the cable puts a restriction on the height to which the coil can be permitted to rise in the tanks. Therefore, for repeatered type-*D* cable, the *Monarch's* capacity is cut back to about 1,600 nautical miles.

Types *A* and *B* cable, used in shallower water, are considerably larger and heavier than type-*D* and, consequently, less of these can be carried.

The ideal laying program would have involved one continuous passage across the North Atlantic from cable station to cable station. However, this would have required carrying over 1,900 miles of ca-

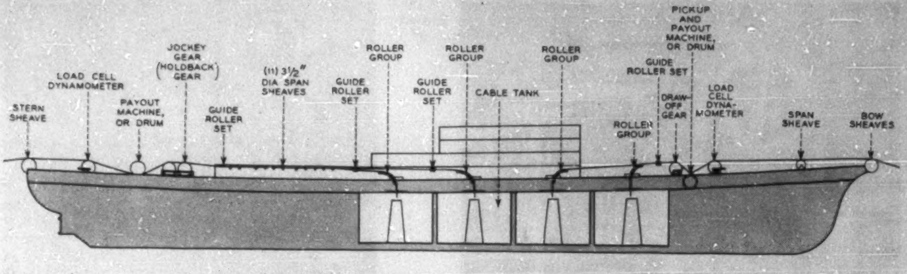


Fig. 10. Line schematic of after-cable gear on Monarch

ble, including about 300 miles of type *A* and something less than 50 miles of type *B*. Such an amount of cable would greatly exceed the ship's capacity.

Each cable was therefore laid in three sections. The no. 1 cable (southernmost), which transmits from west to east, was laid in the following sequence: Clarenville to the mouth of Trinity Bay, a distance of 200 miles; thence about 1,250 miles to Rockall Bank (a submerged plateau); and finally the remaining 500 miles from Rockall Bank to Oban. The no. 2 cable followed the opposite sequence, starting at Oban and proceeding in three sections of 500 miles, 1,250 miles, and 200 miles to the terminal at Clarenville. Shore ends, about 1/2 mile long at Clarenville and 2 miles at Oban, were prepared and put in place in advance of the time when they would be needed. At each intermediate point, the cable was "buoyed off" with a mushroom anchor, connecting lines, and a surface buoy of size appropriate for the water depth.

The mileages indicated are actual cable lengths. They exceed the geographical distances between the points involved because of the slack allowance which experience has shown to be necessary to assure reasonable conformance of the cable with the contour of the ocean bed. Normally, about 5-per-cent slack is considered desirable in deep water, with the allowance decreasing in steps to zero in shallow water.

All available information indicated that the most favorable weather conditions in the North Atlantic could be anticipated in the period of May through August. Before May, ice could be expected along the western sections of the route and after August, hurricanes were likely, and later the winter storms.

Laying of the no. 1 cable was started June 28, 1955, and completed September 26. The actual laying period took but 24 days in this interval, the remainder of the time being spent in transit and in reloading. The no. 2 cable was started June 4, 1956, and completed August 14. About 16 laying days were involved.

The routine aboard ship during laying consisted in passing out cable at the rate of 6 to 7 knots for a repeater section length of a little over 37 nautical miles, then slowing down to about 3 knots as the repeater passed through the cable machinery and overboard, then renewing speed again. During all of this period, the testing crews, both on shipboard and at the transmitting cable station, were busy measuring, recording data, and planning the equalization trimming. At a point shortly after the passage of the next-to-last repeater in an ocean block, special measures were required for the equalization program. From that point until the joints associated with the connection to the following ocean block had been completed, the speed was reduced to 5 knots. The need for this arose from the following considerations.

Stopping of the ship in deep water introduces serious possibility of formation of kinks in the cable, and is to be avoided at all costs. To permit continuous laying, it was necessary to determine the amount of cable needed for equalization, to measure out this cable, and to complete the splices before reaching the end of the block being laid.

The addition of an equalizer at the end of the block requires two joints and armor splices. Preparing the cable ends, brazing together the center conductor and associated tapes, injection-molding the polyethylene around the center conductor, replacing and overlaying the armor wires, and binding the splice, consumes 6 to 7 hours for a single splice, and 8 to 9 hours for two splices when overlapping of operations is practical. An allowance of 3 hours is considered necessary for remolding in event of a defective joint (disclosed by X-ray inspection). The time allowance required to complete the splicing of ocean blocks is therefore 9 to 12 hours. About $1^{1}/_{2}$ hours are needed to carry out the extrapolation, make the equalization decision and turn over cable to the cutting point. During the interval between the cable cut and the completion of the joint, the ship's speed was maintained at 5 knots so as to minimize the distance over which extrapolation of equalization data had to be extended. Even so, the final extrapolation covered the last 60 miles to 75 miles of each ocean block.

During the jointing intervals, system power was turned down to avoid any hazard to the members of the jointing crew. It was restored as soon as the moldings had been X-rayed and the outer or return tapes had been brazed. These activities were so timed that in almost every case the system was powered as each repeater went overboard.

Clarenville–Sydney Mines Link

Route Selection

Clarenville having been selected as the site of the cable terminal station on the west end of the ocean crossing, it was necessary to consider how the system was to be extended to Nova Scotia for connection with the North American continental network.

A number of alternatives were possible, as described in the following.

Alternative 1 contemplated radio relay across Newfoundland to Port aux Basques, and thence across the Cabot Strait. However, a survey revealed that maintenance access to suitable sites would be most difficult, particularly in winter, and primary power was not obtainable.

Alternative 2 involved a poor submarine route around the Avalon Peninsula to possibly Halifax, Nova Scotia. The length of the sea cable would be about 600 nautical miles. It would be necessary to cross many working telegraph cables (see Fig. 3). Trawler damage could be expected as the cable would need to traverse known trawling areas, and during the winter months any repairs would be costly and prolonged. Also, it was not desired to lay another cable out of Trinity Bay as the route might be wanted for a second transoceanic cable at some future time.

Alternative 3 involved a submarine cable from Clarenville out through the North West Arm to Rantem at the head of Trinity Bay, a short land cable across the isthmus, and thence either a submarine cable direct to Sydney, Nova Scotia, or a land crossing of the Burin Peninsula at Garnish and thence by submarine cable to Sydney. This route involved three open-sea sections with one or two land sections. There was rather limited space for a cable in Placentia Bay and the bottom was uneven and rocky. Existing cables laid around the Burin Peninsula have had interruptions which indicated an unsuitable bottom, and fishing trawlers had been seen in the vicinity recently.

Alternative 4 also involved a cable overland, from Clarenville to Terrenceville at the head of Fortune Bay, there to join a direct submarine cable to Sydney Mines. Three short underwater sections would be involved in the Clarenville-Terrenceville link, but these could be in shallow water, out of harm's way. The main submarine route from Terrenceville to Sydney Mines would be clear of other cables and would avoid trawling areas and anchorages, a not inconsiderable achievement in view of the congestion of submarine cables and the fishing activity around the southeast corner of Newfoundland. Further, a good landing site in Nova Scotia was available near Sydney Mines on property owned by Eastern Telephone and Telegraph Company.

After due consideration, alternative 4 was chosen as being the most satisfactory from all aspects, and the final route is shown in Fig. 3. This route is considered to be most likely to have a good life history. While it would have been possible to have one continuous land cable between Clarenville and Terrenceville, the three short underwater sections saved a considerable amount of trenching without adding undue hazard to the system.

CLARENVILLE-TERRENCEVILLE ROUTE

Routing the Post Office single-cable system overland from Clarenville to Terrenceville having been decided upon, a number of other matters required decision. The first was the type of cable to be employed for this section. Several alternatives were considered, bearing in mind factors such as the type of terrain, access, availability of primary power, possible future expansion of capacity and, of course, interference from static and radio frequency pickup. The advantages of using standard solid dielectric coaxial ocean cable with submarine-type repeaters were judged to outweigh all other considerations, and left only the problem of shielding from interference.

Up to this time, shore ends of submarine cables used by the Post Office were shielded for about a quarter of a mile from shore by a lead sheath insulated from the return tapes of the coaxial by a polyethylene barrier. Experience indicated that such shielding might not be adequate over a long distance on land. The question was resolved by the addition of iron shielding tapes and a plastic jacket to standard submarine cable. The structure is described elsewhere.[5]

Through the use of this robust, wire-armored cable and two steel-housed submarine repeaters, no limitations from the noise pickup angle were placed on the detailed route selection for the overland section. The first thought was to try to proceed directly across country from Clarenville to Pipers Hole River, saving at least 10 miles over a route which followed the road, and on which advantage might be taken of quite long water stretches into which the cable could be dropped. Black River Pond, for instance, is $4^1/_2$ miles long. This proposal was abandoned after surveys, because of the very rocky nature of the country and the difficulty of access both for construction and any subsequent maintenance, and it was decided to follow the general course of the roads.

It was possible to avoid trenching in the rocky, precipitous cliff country from Clarenville to Adeytown by laying about 6 miles of cable in the water of Northwest Arm. Similar considerations dictated the choice of 2 miles of cable in the sea across Southwest Arm. Thence the route followed the road, at a distance ranging from 250 yards to more than a mile, as far as Placentia Bay, taking advantage of the larger ponds, where possible, to avoid trenching.

Reaching the 800-foot-high ground beyond the Pipers Hole River from the north of Black River proved quite difficult. Here the road is carved out of the foot of the cliffs as far as Swift Current and the country behind is solid rock. Plans exist for a hydroelectric project involving dams in Pipers Hole River just north of the road crossing and it naturally would not be desirable to bury a cable in such a locality. The river estuary itself passing by Swift Current presents only a narrow 6-foot deep navigable channel at low water but it was decided that this could be used for some 6 miles by employing a barge and a shallow draft tug for the laying.

To locate a suitable route out of the basin up through wooded gorges to the top took about a week of very rough work. Thereafter, all was "plain sailing;" advantage was taken of ponds such as Long Pond (4 miles) and Sock Pond (3 miles) until the route arrived within 6 miles of Terrenceville.

Here it was reluctantly decided to bury the cable in a deep trench on the inner side of the road, as the other side falls sharply to the sea. The desire to avoid roads was due to their instability and the methods used for construction and repair. This road is dirt only, with no foundations, and in this particular area, sections have been known to slide away into the river bed below. The final length of cable laid was just short of 55 nautical miles.

CABOT STRAIT LAYING

Coaxial submarine cables in which rigid repeaters are inserted cannot be laid with the existing cable-laying machinery except by stopping the ship, removing the turns of cable from the drum, and then passing the repeater by the drum and restoring the turns. Special equipment is also needed for launching the repeater over the bow sheaves.

SHIP MODIFICATIONS

The following equipment was installed on the *Monarch* for the laying of cables carrying rigid repeaters: A gantry over the bow baulks, consisting of a 22-foot steel beam projecting 6 feet beyond the sheaves, was installed for handling repeaters at the bow. This gantry was fitted with an electrically operated traveling hoist for lifting the repeaters over the bow sheaves and lowering them into the sea. A stand-by hand-operated lifting block and traveller were provided to guard against failure of the power point.

A rubber-tired, steerable steel dolly (or trolley) was developed from the chassis of a small car to transport the repeaters from the cable tank hatches to the bow sheaves. These repeaters weigh about 1,200 pounds.

Storage racks were built up from steel sections provided with shaped, rubber-lined wood blocks and were fitted at each cable tank hatch on the shelter deck. Each rack held four repeaters in double tiers. A hand-operated lift was furnished for moving the repeaters from the storage racks to the dolly.

A special quick-release grip was furnished for use when lifting the repeaters by the electric hoist on the bow gantry. Deflection plates were also fitted on the fore deck around dynamometers and hatches to avoid their fouling the dolly.

LAUNCHING RIGID REPEATERS

The rigid repeaters were stowed in their racks in the order of their laying. The bights of the cable attached to the ends of the repeaters were brought up the sides of the cable tanks, secured along the arms of the crinoline and up the sides of the hatch coamings to the deck, clear of the running length of cable and from where the repeater could be drawn forward along the deck on the dolly.

When the time came for a repeater to be laid, speed was reduced and the ship finally stopped head to wind. A 6-by-3 compound rope from the starboard cable drum was secured to the cable just abaft the bow baulks. Sufficient cable was then paid out until the tension was taken up by this rope. The turns of the running cable were then removed from the port cable drum and the resulting slack cable worked overboard by paying out the starboard drum rope which was holding the tension. When the excess cable had been cleared from the deck, the repeater on its dolly was carefully hauled along the fore deck to the traveling hoist of the overhead gantry. Cable was then drawn from the tank so that the turns could be re-formed on deck and replaced on the port cable drum. The repeater was lifted from the trolley and traversed outboard as soon as it was high enough to clear the bow baulks. It was then lowered to the water's edge and when the tension had again been taken by the cable, the quick release grip was slipped and the starboard drum rope cut. Paying out was then resumed.

LAYING PROGRAM

On February 1, 1956, the *Monarch*, having returned to Ocean Works, Erith, after refitting, commenced loading the various sections of cable to be used for the Terrenceville–Sydney Mines route. The sections were all carefully tested and measured in the Works before loading. The cable ends were clearly marked and dogged together by a length of rope which was not removed until the repeater had been jointed into its connecting sections of cable.

Loading of the cable and splicing in of repeaters was finished by April 10 and the system tested and checked. The *Monarch* sailed for Sydney Mines on April 18 and arrived there April 30. The cable station is situated about $1^1/_2$ miles inland from the shore, with a small lake intervening. A length of type-*B*, insulated-outer-conductor, lead-covered cable had previously been laid from the station across the lake to a narrow strip of land which separates it from the sea. The joint to the main cable was to be made on this strip. Two medium-sized shore-based motor boats were used to tow the end of the double-armored section of cable from the *Monarch* to the shore. During this journey the cable was supported by empty oil drums at close intervals. When the motor boats had reached shoal water, the end of the cable was secured to a landing line and two tractors took over the hauling.

When enough cable was on shore to make the joint and the splice, the barrels were cut away and the *Monarch* weighed anchor and paid out this section of double-armored cable on the agreed route and buoyed off the end. She then steamed over the proposed track to Terrenceville, taking soundings and sea bottom temperatures as required, and anchoring off Terrenceville on May 3.

Preparations for landing the end were at once put in hand and the ship's motor launches towed the end of the cable towards the cable landing, the cable again being supported by empty oil drums. This end was jointed and spliced to a piece of cable which had been laid previously from the Terrenceville cable hut to a sand split which juts across the head of Fortune Bay, about a mile away.

Upon completion of the splice, over-all tests were made from the ship to the Terrenceville cable hut and, all being well, paying out toward the buoyed end off Sydney Mines was begun on May 4. The first repeater went over about 2 hours after the start of laying and the others followed at approximately $4^3/_4$-hour intervals.

On May 7 the cable buoy on the Sydney Mines end was recovered and the end hove inward. After tests in both directions, the final joint and splice were made. This operation was completed on May 9 and on receipt of a signal that all was well the *Monarch* proceeded into harbor at Sydney to land testing equipment, a spare equalizer, and other equipment.

EQUALIZATION AND TESTING

The cable had been loaded into the ship in repeater section lengths, so cut that when laid at estimated mean annual sea temperature, the expected attenuation would be 60 decibels at 552 kc. A correction for the change in attenuation of the cable when coiled in the factory tanks and when laid in about 100 fathoms had been determined from tests on two 10-mile lengths of cable, laid off the Island of Skye. The correction amounted to a decrease in attenuation, when the cable was laid, of 1.42 per cent. This was essentially an empirical result, and as the mechanism of the change was not fully understood, a possible further inaccuracy of equalization might arise.

Sea-bottom temperatures along the route were obtained from information supplied by the Fisheries Research Board of Canada but, unfortunately, this information was rather meager and varied considerably with locality.

Since the cable equalization built into the repeater differed appreciably from the final determination on laid cable, it was found necessary at a comparatively late stage to introduce an undersea equalizer into the center of the sea section. This was intended to eliminate a flat peak of loss of 3.5 decibels, expected at about 100 kc. So that the last repeater should not lie too near the beach at Sydney Mines, a network simulating 9 miles of cable was also inserted in the undersea equalizer.

The repeaters were spliced into the cable lengths on board the *Monarch* and tests were made at every stage of the build-up of the system. The equalizer was permanently jointed to the first half section of seven repeaters and left with an excess length of tail which could be cut as desired during the laying operation to further improve the equalization. The first and second halves of the system were temporarily connected through power separation filters so that the whole system could be energized just prior to laying.

The test routine included attenuation measurement at five frequencies in each direction, of transmission, noise, pulse and loop-gain, supervisory measurements, d-c and insulation resistance, and capacitance. The *Monarch* test room therefore contained two sets of terminal equipments similar to those installed at Clarenville and at Sydney Mines.

It was decided to energize the system continuously during the laying except for the few hours when power had to be removed to make the equalizer splice. This enabled a continuous-order (speaker) circuit to be operated over the cable, and minimized the number of energizing and warm-up periods. The only disadvantage, considered to be slight, was the necessary omission of insulation resistance and capacitance measurements during laying, except in the course of the equalizer-splicing operation.

The plan was to lay from Terrenceville in the direction of the high-frequency band and to test the system to the *Monarch* during the laying from this shore station. The overland section between Clarenville and Terrenceville, which contained two repeaters, was connected on with appropriate equalization after the submarine section had been satisfactorily completed and tested.

At Terrenceville, after the cable end had been taken ashore and the beach joint completed, the system was energized from the *Monarch* with a d-c power ground at Terrenceville for the necessary 4-hour minimum warming-up period. The first set of routine measurements of the laying operation was then carried out. Thereafter, a complete set of measurements on the Terrenceville half of the system was made after every 10 miles of cable laid. An occasional check set of measurements was also made on the Sydney Mines half of the system in the tanks.

The primary object of these tests was to determine what length of cable should be inserted between the equalizer and the eighth undersea repeater to obtain the optimum system. In practice this resulted in arranging for a length of cable so that the output level of the eighth repeater at 522 kc should be equal to that at the output of the first repeater at the assumed mean annual temperature of 35.1 degrees Fahrenheit.

The estimated length of cable required for this purpose was plotted after each measurement. It became evident soon after laying the fifth repeater (94.6 nautical miles of cable laid) that the linear relation obtained could be extrapolated with adequate accuracy to specify safely a length of 6.06 nautical miles of cable between the equalizer and the adjacent repeater.

This decision on length was taken and, after removing power, the equalizer was accordingly jointed in, the operation being completed before it was necessary to pay out the splice. During this period, capacitance and conductor and insulation resistances were checked on each half. During the laying of the second half, measurements were made as for the first half. On arrival at the buoyed shore end, a final complete set of measurements was made and these, suitably corrected for the shore and length, were transmitted to Sydney Mines so that the first measurements from Sydney Mines to Clarenville could be checked with those obtained on the ship.

Sidelights

Weather

Weather is the undetermined element in cable-laying and repair activities. With few exceptions the transatlantic project was blessed with remarkably good weather. The exceptions were, however, noteworthy.

Heavy snow squalls were encountered off Terrenceville during the operations in the vicinity. At Rockall Bank on the first lay one heavy storm came up as the last repeater in the 1,250-mile section was going over, and made this launching and the subsequent buoying of the end very difficult operations. Both were accomplished successfully as a result of the superb seamanship of the *Monarch's* commander, Capt. J. P. F. Betson, and his officers and crew.

A second, and worse storm was encountered upon the return to Rockall. This was a manifestation of hurricane Ione, with wind velocities of over 100 mph and very high seas. The ship had given up searching for the buoy (later reported drifting more than 500 miles away off the Faeroe Islands) and was grappling for the cable, when the storm hit. Fortunately, the cable had not yet been found, so the ship could head into the wind and ride it out. She was driven many miles off course in the process, and the seas will be long and vividly remembered by all present. Incidentally, the cable was picked up shortly after the storm had moderated.

Generally speaking, the effect of the weather on the engineering supernumeraries on board was not severe, although the *Monarch's* stock of dramamine was somewhat depleted by the end of the project.

Miscellaneous Events

At the start of the first transatlantic lay, several icebergs were encountered. A small one at the mouth of Random Sound lay in the path of the cable and therefore caused an involuntary, though minor, revision of the route. The others, beyond the mouth of Trinity Bay were larger, but also farther away. Whales and grampuses became common sights, though much photographic film was expended at first by the uninitiated. An occasional bird rested on the ship far from land, obviously exhausted from its long and presumably unintended journey.

Progress Bulletins

Daily progress bulletins were radioed to headquarters of all partners during the laying. In addition, because a telephone cable system differs considerably from submarine telegraph cables, the officers and crew were briefed by the engineering personnel as to the repeater structure, the need for equalization, and the general objectives of the venture. This proved to be a very profitable move indeed, for the co-operation of all hands was everything that could be wished. As a follow-up, daily performance bulletins were posted in strategic parts of the ship so that everyone, no matter what his duties, could know just how the evolving system was performing with respect to objectives.

Cable Order Circuit

One-way conversation from shore to ship over the cable was possible all the time the repeaters were energized. This was a source of very great satisfaction to the shipboard test crew, as it was concrete evidence that the cable was working, and working well. When power was turned down, the recourse to radio telephone provided a comparison which generally left no doubt as to the future value of the cable.

When the final splice was slipped into the water of Clarenville Harbor, on August 14, 1956, there was completed a venture quite unique in the annals of submarine cable laying. And in the laying perhaps more than in any other phase of the transatlantic project the successful conclusion provided evidence of the friendly and harmonious relationships among the different organizations and nationalities involved, and of the close co-ordination of their efforts.

References

1. Repeater Design for the North Atlantic Link, T. F. Gleichmann, A. H. Lince, M. C. Wooley, F. J. Braga. Page 865 of this issue.

2. System Design for the North Atlantic Link, H. A. Lewis, R. S. Tucker, G. H. Lovell, J. M. Fraser. Page 850 of this issue.

3. Cable Design and Manufacture for the Transatlantic Submarine Cable System, A. W. Lebert, H. B. Fischer, M. C. Biskeborn. Page 907 of this issue.

4. Power-Feed Equipment for the North Atlantic Link, G. W. Meszaros, H. H. Spencer. Page 890 of this issue.

5. System Design for the Newfoundland–Nova Scotia Link, R. J. Halsey, J. F. Bampton. Page 917 of this issue.

(*Late discussion and erratum on page 956.*)

LATE DISCUSSION

The following discussion was received too late to be included with the paper.

Discussion of paper 56-207 by Kan Chen and A. J. Schiewe, published in Communication and Electronics, Sept. 1956, pp. 396–400.

A Single-Transistor Magnetic-Coupled Oscillator

Kan Chen and A. J. Schiewe: The authors wish to thank Mr. Van Allen and Mr. Mortimer for their discussions of this paper. In particular, Mr. Mortimer is to be commended for his fine work which is recorded in his Naval Research Laboratory report. His efforts appear to be very nearly parallel to those of the authors in regard to the single-transistor oscillator.

The ideas presented in our paper are intended to give a complete explanation of the operation of Van Allen's circuitry. This, then, forms a basis for the logical development of the single-transistor oscillator. It is felt that past reports, although possibly dealing with essentially the same circuitry, have neglected such an approach.

Chronologically, our work, just as Mr. Mortimer's, was preceded by a much wider effort in solid-state device research which was being carried on by Westinghouse Electric Company. A review of records indicates that a Westinghouse materials engineering department engineering report, from which our paper evolved, was issued on February 24, 1955. The publication of our work in its present form was delayed by the fact that one of the authors is now at Purdue University.

ERRATUM

"Dynamax, a New Crystal and Domain-Oriented Magnetic Core Material" by G. H. Howe, published in Communication and Electronics, November 1956, pages 548–51, closure by G. H. Howe.

An error was made in reproducing Figs. 4 and 5. They should be transposed, so that the right-hand chart appears over the caption "Hysteresis loops of 2-mil 65 permalloy," and the left-hand chart appears over the caption "Hysteresis loops of 2-mil Dynamax."